PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 14 May 15, 1928 Number 5

ON THE QUANTUM THEORY OF THE AUTOELECTRIC FIELD CURRENTS

By J. R. Oppenheimer*

Norman Bridge Laboratory, Pasadena

Communicated March 28, 1928

By the application of a sufficiently intense electric field a Bohr atom could be rendered unstable. For when the drop in potential across the electronic orbit reaches a value of the order of magnitude of the ionizing potential of the atom (10^9 volts/cm.), the electron, instead of remaining in the neighborhood of the nucleus will fall down the hill of potential energy, and the atom will be dissociated. This dissociation is explosive in character. For there is a critical field strength, below which the atom remains stable indefinitely and above which it dissociates in a time of the order of the orbital periods of the atom. The characteristic for the autoelectric current should accordingly show abrupt discontinuities.

If one examines the same problem with the quantum mechanics, he finds that these abrupt changes disappear: Any field, no matter how weak, will in time dissociate an atom. This is essentially a consequence of the fact that the motion of the electron is no longer absolutely restricted to a region of the dimensions of the Bohr orbit; it will now occasionally, though not very often, be found at points much further from the nucleus; and the further it is, the smaller will be the field required to insure that it does not return to the nucleus. Since the probability that an electron be at a distance R from the nucleus falls off exponentially with increasing R, the rate at which the field ionizes the atom may be expected to decrease rapidly when the field strength is decreased.

These considerations may be made precise. For, by a slight extension of the perturbation theory of the quantum mechanics, one may find the rate at which the field induces transitions; and the particular transition which is responsible for the effect is the quantum jump from the normal state of the atom to a state of the same energy, in which electron and nucleus are falling apart under the influence of the field. This rate turns out[1] to be, for a hydrogenic atom with ionizing potential W,

$$0.02\, h/m\, a^{-5/4}\, \alpha^{1/4} \exp\{-\tfrac{4}{3}\, a^{-3}\alpha^{-1}\} \qquad \begin{matrix} a = h/2\pi(2mW)^{-1/2} \\ \alpha = 8\pi^2 meF/h^2 \end{matrix} \qquad (1)$$

per unit time; here m and $-e$ are the mass and charge of the electron. If the wave functions of the atom are not quite hydrogenic in character, the exponential is not essentially affected, but the numerical factor may vary by a factor, perhaps, of two.

From (1) it follows that the characteristics of the field currents should satisfy

$$ln\, i - \tfrac{1}{4}\, ln\, F = C/F. \qquad (2)$$

The reversible curves published two years ago by Millikan and Eyring[2] do, in fact, satisfy this condition very satisfactorily. And the formula (2) is essentially that found empirically and independently by Millikan and Lauritsen;[3] it has been shown by them to hold with great exactitude for all the reversible currents investigated.

With W in volts, the theoretical value of C in volts per cm. is $-10^8\, W^{3/2}$. The observed slope of the characteristic should thus make possible, provided F may be correctly estimated, the determination of the ionizing potential for electrons in metal: the work function. The values of F, however, computed from the geometrical dimensions of the wire, give for W values of a few tenths of a volt. From this we may conclude that the autoelectrons do not come uniformly from the surface of the wire, but only from certain favored points where the field strengths are abnormally high. Now this is just what is found experimentally. For the wire (after systematic out-gassing) may be subjected to fields of four million volts per centimeter without drawing a current measurable on the galvanometer;[4] but this condition may suddenly be changed,[3] so that the current jumps immediately to a large value, and one (or more) bright spots appear on the anode and the walls of the tube. If after such an event the wire is examined[4a] microscopically, it is found that at appropriate positions there are small craters, and that these are surrounded by protuberances with very small radius of curvature; and the dimensions of these points, which should be responsible for the current, vary from a tenth to a fortieth of those of the wire and thus give work functions of the order of a few volts. Thus a typical crater has a radius of curvature of 2×10^{-5} cm. and gives $W = 4.7$ volts. Diffraction, however, makes the precise estimation of the radii of curvature of the points impossible.

The formula (1) gives in amperes for the current per series electron:

$$i = 3 \times 10^{-8}\, F^{1/4}\, W^{1/4} \exp -(10^8\, W^{3/2}\, F^{-1}). \qquad (3)$$

It is thus possible, from the value of W and F, to compute the number of atoms taking part in the effect. This computation cannot be made very

precise, however, because small changes in the exponential mask large changes in the number of atoms, and because, further, the value of the constant factor is not quite certain. But all the values so far obtained show that the number of atoms taking part lies between 10^{-4} and 10^{-6} of the total number of surface atoms. This explains the spotted appearance of the anode, and agrees reasonably with the dimensions of the points as observed, which give an area about 10^{-5} that of the wire.

The field currents are not absolutely independent of temperature. Empirically the characteristic curves show two distinct slopes for constant temperature and varying field; these correspond respectively to the thermionic current as modified by the field (Schottky), and to the field current as modified by the temperature; because of the steepness of the exponentials the intermediate region is not susceptible of experimental resolution. The former region is given qualitatively by Schottky's theory of the reduced image force. In the latter region one finds the law that the relative change in current induced by a change of temperature is nearly independent of the strength of the field. This law follows readily from the theory of field currents,[1] if one assumes the existence of one or more excited electronic states with work functions somewhat less than W.

Since the formula (1) was published, two theoretical treatments of the autoelectric effect have been given. The former of these, that of Houston,[5] ascribes the effect to the reduction of the work function of the metal by the field, and the internal partial pressure which the conduction electrons acquire on the Fermi statistics. This seems, however, not to lead to a satisfactory quantitative account of the experiments; for this purpose, the atom furnishes a better model than the Fermi gas. In the second paper[6] the effect is obtained, as in the present theory, from the fact that the wave functions for the electrons do not vanish completely outside the surface of the metal; but the method of computing the current is inconsistent with the quantum mechanics and involves undetermined constants which should in fact be determinate. The characteristic obtained by this calculation also fails to agree with experiment, because the binding force which holds the electron in the metal is taken to be the image force instead of the coulomb attraction of the ions, as in the present paper.

* NATIONAL RESEARCH FELLOW.

[1] J. R. Oppenheimer, *Phys. Rev.*, **31**, 80 (1928).

[2] R. A. Millikan and C. Eyring, *Ibid.*, **27**, 55 (1926).

[3] R. A. Millikan and C. Lauritsen, *Proc. Nat. Acad. Sci.*, **14**, 45 (1928).

[4] R. A. Millikan and B. E. Shackelford, *Phys. Rev.*, **15**, 239 (1920).

[4a] I am indebted to Dr. Lauritsen for an opportunity to see his photographs of the craters.

[5] V. Houston, *Zs. Phys.*, **47**, 33 (1928).

[6] O. W. Richardson, *Roy. Soc. Proc.*, **A667**, 719 (1928).

THE FERMI STATISTICAL POSTULATE; EXAMINATION OF THE EVIDENCE IN ITS FAVOR

By Edwin H. Hall

Jefferson Physical Laboratory, Harvard University

Communicated April 14, 1928

My interest in the Fermi postulate is due mainly to the use which Sommerfeld makes of it in his theory of metallic conduction. The Fermi statistics is opposed by the Bose-Einstein statistics, and it seems desirable to indicate the essential difference between the two.

In 1924 Bose,[1] undertaking to deal statistically with "light quanta," hit upon a method of using generalized coördinates which enabled him to derive the Planck law of black-body radiation without making use, as others had done, of conceptions taken from the classical mechanics. Einstein,[2] adopting the Bose statistical method, applied it to the discussion of quantization of energy in an ideal monatomic gas, and the consequent degeneration [Entartung] of the gas—that is, the degree of its departure from the classical gas-laws. He reached the following conclusion: "The mean energy of the gas molecules (as well as the pressure) at the temperature is therefore always less than the classical value."

In 1926 Fermi[3] attacked the same problem that Einstein had discussed. He used the same general statistical *method* that Bose had invented but introduced in its application a certain limiting assumption, with the following explanation: "Now a short time ago the rule was proposed by Pauli[4] . . . that when in an atom an electron exists the quantum numbers of which (the magnetic quantum number included) have definite values, there can be in the atom no other electron whose path is characterized by the same numbers. In other words, a quantum orbit (in an external magnetic field) is already completely occupied by a single electron." "Since this rule of Pauli has proved extremely fruitful in the explanation of spectroscopic facts, we propose to inquire whether it is not also of some use for the problem of the quantization of ideal gases." "We shall therefore in what follows assume that at most one molecule with given quantum numbers can exist in our gas; as quantum numbers not only those here come into consideration which belong to the internal motions of the molecule but also those which determine its motion of translation."

It will, I think, be generally admitted that, granting the truth of the Pauli principle as applied to the relatively few, evidently related electrons within a single atom, the extension of this principle to all the particles in even one gram-molecule, 6×10^{23} particles, is on *a priori* grounds an exceedingly doubtful step, especially in view of the fact that Fermi makes no attempt to give a physical explanation of the supposed in-

fluence of the molecules upon each other. Sommerfeld in the first published sketch[5] of his conduction theory calls this assumption bold, admits that the physical operation of the "principle" is entirely a mystery [völlig undurchsichtig], but says that "the consequences of the postulate must be thought out."

One of these consequences, and one which Fermi was avowedly hoping to find, was a greater degree of "degeneration" of the ideal gas at low temperatures than other discussions of the matter had indicated. Another consequence is stated by Fermi as follows, just after his equation (34):

"The pressure is therefore greater than according to the classical equation of condition ($p = nkT$). For an ideal gas with the atomic weight of helium at $T = 5°$ and a pressure of 10 atmospheres the difference amounts to 15 per cent." After equation (36) he remarks: "One can see by this that the degeneration has a zero-point pressure and a zero-point energy as a consequence."

Thus the Fermi statistics, with the assumption mentioned, is in direct opposition to the Bose-Einstein statistics regarding the direction of departure of ideal gases from the classical laws at low temperatures.

Direct experiment upon gases at low temperatures cannot at present decide between the claims of the two theories. Preference for one theory or the other must be justified on other grounds, and in this connection we come upon the name of Dirac. In fact, Sommerfeld brackets this name with that of Fermi in his reference to the statistics, or to the assumption, in question. Dirac's paper[6] contains in §4 a "Theory of the Ideal Gas," applying thereto the method of wave-mechanics, to which the term *eigenfunction* belongs. I make no pretense of having sounded the depths of its mathematical processes. But the logic of the following passage, the most significant one in the paper for our present discussion, is not difficult to follow, though I think the author must have been surprised to find how much importance has been attached to its final sentence: "We must now make the assumption that all stationary states of the assembly (each represented by one eigenfunction) have the same *a priori* probability. If now we adopt the solution of the problem that involves symmetrical eigenfunctions, we should find that all values for the number of molecules associated with any wave have the same *a priori* probability, which gives the Einstein-Bose statistical mechanics. On the other hand, we should obtain a different statistical mechanics if we adopted the solution with anti-symmetrical eigenfunctions, as we should then have either 0 or 1 molecules associated with each wave. The solution with symmetrical eigenfunctions must be the correct one when applied to light quanta, since it is known that the Einstein-Bose statistical mechanics leads to Planck's law of black-body radiation. The solution with anti-symmetrical functions, though, is probably the correct one for

gas molecules, since it is known to be the correct one for electrons in an atom, and one would expect molecules to resemble electrons more closely than light-quanta."

This is the whole, I believe, of Dirac's argument in favor of extending the Pauli "equivalence principle," or exclusion principle, from the few electrons of a single atom to the countless molecules of any body of gas no matter how great. Is this argument conclusive? Does it, for one who ponders, even raise a presumption in favor of the proposition advanced so casually by the author? It may be that a molecule resembles an electron more than it resembles a light-quantum; but is it true that an extended gas, with countless trillions of molecules, resembles the group of electrons within a single atom more closely than it resembles an extended radiation space filled with light quanta? Sommerfeld himself, though in his *Naturwissenschaften* paper of last October he seems disposed to accept the Fermi proposition without qualification, merely calling it "bold" and its physical cause "völling undurchsichtig," in his more extended, recent statement[7] of his conduction theory says, on page 5, "Whether the extension, proposed by Fermi, of his statistics to the ordinary, electrically neutral gas is legitimate appears doubtful."

This statement casts as much doubt upon the authority of Dirac as upon that of Fermi, and it thus seems to weaken the very foundation upon which Sommerfeld has constructed his whole electron-theory of metals. There remains, however, Pauli as a support, whose admittedly successful application of the "equivalence principle" to certain atoms has been supplemented by whatever measure of success has attended his application of the Fermi statistics to the problem of paramagnetism.

How great has this success been? Sommerfeld begins the recently published account of his theory with this statement: "W. Pauli[8] has shown that the electron-gas of the Drude theory of metals is highly degenerated in the sense of the Fermi statistics, and could from this explain quantitatively the hitherto puzzling, of-temperature-independent, paramagnetism of the alkalies." This seems to me to go somewhat beyond the claims made by Pauli himself in the paper referred to.

Pauli in the early part of his paper discusses at considerable length the question whether the Bose-Einstein or the Fermi statistics should be assumed to hold among the free conduction-electrons of a metal, which he takes to be identical with the valence electrons. He refers to papers[9] by Heisenberg, who applies quantum mechanics to "systems which consist of several exactly similar particles" and finds that more than one solution will satisfy the mathematical conditions of his problem. I now quote Pauli: "By only one of these solutions . . . is the 'equivalence rule' fulfilled, and it can be regarded as empirically certain that only this solution is realized in nature." This last statement appears very sweep-

ing, but what immediately follows shows that what Pauli has in mind as "empirically certain" is merely the success of his equivalence principle *within the atom*.[10] For he continues thus: "(Especially ni He, through rejection of the second solution, is explanation found of the fact that the total magnetic moment, in the normal condition, of the He atom vanishes.) It is thus proved that the quantum-mechanics is not in conflict with the equivalence rule. On the other hand, for this special selection among the solutions of the quantum-mechanics equations and so, too, for the equivalence rule, a satisfactory theoretical ground is still lacking."

Pauli goes on for about a page further to argue the question as to the relative merits, for his problem, of the two opposed statistics, the Bose-Einstein system and the Fermi system, without finding any conclusive answer. But he finally decides to try the latter.

He now, page 85, states his problem thus:

"In what follows it will be especially sought to show how on the ground of this mode of statistics the fact that many metals are diamagnetic or only very weakly paramagnetic can be brought into harmony with the existence of a magnetic moment of the electron. Herein the conduction-electrons in the metals considered will be treated as an ideal gas—that is, the forces which act upon the conduction-electrons will be disregarded. This is certainly a gross approximation and must be replaced by a more exact model of the metals; it is, nevertheless, not improbable that the here-offered attempt at an explanation, at least as regards the qualitative side of the phenomenon, approaches the correct one."

On page 101 he gives a comparison between calculated and observed results, χ_0 being the magnetic susceptibility at absolute zero:

	NA	K	RB	CS
χ_0 calc.	(6.57	5.2	4.88	4.54) 10^{-7}
χ_0 obs.	(5.8	5.1	0.6	-0.5) 10^{-7}

The concluding paragraph of his paper is this:

"More than agreement in order of magnitude is, with the provisional character of the ideal gas model for the conduction-electrons in the metal, not to be expected. Especially for a theory of the gradual going over of the susceptibility, in melting and evaporation of the metal, into the values calculated according to the Curie law (independence of the sodium atoms as to each other), it will be necessary to take account of the forces acting on the valence electrons."

It would certainly be unfair to say that Pauli's application of the Fermi statistics to the problem of paramagnetism has failed. But, on the other hand, one may seriously ask whether its success has been striking enough and extensive enough to establish beyond doubt the essentially improbable and admittedly mysterious fundamental hypothesis of the

Fermi statistics. When we recall the fact Sommerfeld has recently expressed doubt as to the validity of this hypothesis in the case of electrically neutral gases, and the further fact that Einstein has given a statistical treatment of the ideal gas problem which meets the mathematical requirements of the case without using the Fermi hypothesis, we are surely justified in holding that the Fermi statistics is still on trial and that its fate must depend very much on the degree of success which may attend the effort of Sommerfeld to base upon it an electron-theory of metals.

[1] *Zs. Phys.*, **26**, 3, pp. 178–181.
[2] Berlin, *Berichte*, **1924–25**, pp. 261–267.
[3] "Zur Quantelung des idealen einatomigen Gases," *Zs. Phys.*, **36**, 902.
[4] W. Pauli, Jr., *Ibid.*, **31**, 765 (1925).
[5] *Naturwissenschaften*, Oct. 14, **1927**, p. 63.
[6] *Proc. Roy. Soc., London*, **A112**, 1926.
[7] *Zs. Phys.*, **47**, pp. 1–32.
[8] "Uber Gasentartung und Paramagnetismus," *Zs. Phys.*, **41**, 81 (1927).
[9] *Zs. Phys.*, **38**, 411 (1926); **39**, 499 (1926).
[10] No one, I believe, extends the original Pauli exclusion principle from one atom to another. That is, no one asserts that two electrons in two different atoms, even adjacent atoms, may not have equivalent quantum numbers.

SOMMERFELD'S ELECTRON-THEORY OF METALS

By Edwin H. Hall

Jefferson Physical Laboratory, Harvard University

Communicated April 14, 1928

Sommerfeld's theory is based upon the Fermi statistical mechanics and must, I believe, stand or fall with the fundamental postulate of Fermi's argument, the extension of Pauli's "equivalence principle" to all the molecules of a monatomic ideal gas, no matter how numerous these molecules may be. I have dealt with this postulate especially in another paper printed in this issue of these Proceedings.

The paper or papers which I shall here discuss may be found in the *Zeitschrift für Physik*, volume **47**, February, 1928, pp. 1–60.

Pressure and Kinetic Energy of the "Electron-Gas."—Taking silver as an example and assuming the particles of his conduction "electron-gas" to be equally numerous with the atoms of the metal, Sommerfeld, on page 15, finds as the pressure of this gas at the absolute zero of temperature, 200,000 atmospheres. The corresponding kinetic energy of the gas per unit volume of the metal is 3×10^{11} ergs, and dividing this amount by 5.9×10^{22}, the number of silver atoms, and so by Sommerfeld's hy-

pothesis the number of gas-electrons, per cu. cm., we get, as the kinetic energy of a single electron at the absolute zero of temperature, 5.1×10^{-12} ergs. This is about 80 times the kinetic energy of an ordinary gas molecule at 300° absolute and about 22 times the kinetic energy of such a molecule at 1100° absolute, a temperature at which thermionic emission of electrons from metals occurs to a marked extent. The significance of these numerical values will presently appear.

Electric Conductivity.—Sommerfeld adopts the Lorentz theory of electric conduction by means of the free electrons, the "mean free path" of the electron among the atoms being conceived of as analogous to the mean free path of the kinetic theory of gases. Other things being equal, conductivity will on this basis be the greater the greater the duration of the free path. In order to account for the conductivity of silver Sommerfeld has to make the time-duration of the free path such that, with the very high average velocity corresponding to the high kinetic energy of the electrons, the mean free path proves to be about 5.2×10^{-6} cm. long, or, as Sommerfeld remarks "some hundreds of times the atomic distance," distance from center to center of neighboring atoms.

Now one of the long-standing difficulties of the "classical" theory of conduction by means of free electrons has been that, if these electrons were to share the heat energy of thermal agitation, their thermal capacity would make the specific heat of metals greater than it is. Sommerfeld's theory gets rid of this difficulty by making the kinetic energy of the conducting electrons nearly independent of temperature. But the apparent magnitude of this difficulty in the classical theory was due to the fact that only very short mean free paths were supposed to be allowable, and so a large number of free electrons was required. If, however, free paths of such length as Sommerfeld uses are permissible, a comparatively small number of free electrons will serve, and so the specific-heat difficulty disappears[1] from the classical theory of free-electron conduction.

In discussing Sommerfeld's treatment of thermionic electron emission I shall refer again to the high kinetic energy which he attributes to his conduction-electrons.

The Lorentz formula for free-electron conductivity, too familiar to need explanation, is

$$\sigma = \frac{4}{3} \frac{e^2 l n}{\sqrt{2\pi m k T}}. \qquad \text{Sommerfeld's (48}a\text{).}$$

Sommerfeld's corresponding formula is

$$\sigma = \frac{8\pi}{3} \frac{e^2 l(\bar{v})}{h} \left(\frac{3n}{8\pi}\right)^{\frac{2}{3}}, \qquad \text{numbered (48}c\text{).}$$

With regard to the latter equation Sommerfeld now remarks: "Char-

acteristic here in contrast with (48*a*) is the occurrence of h and the different dependence on n. The explicit dependence on T has disappeared; moreover, we must take $l(\bar{v})$ [this expression means that the free path l is some function of the velocity, $\bar{v}$ being something near the mean velocity of the electrons] as well as $(\bar{v})$, as in the first approximation independent of the temperature. There lies the most difficult point of the whole theory. In order to explain the empirically known dependence of σ on temperature it would be necessary to push further the conception of the free path, therefore . . . to take account of the heat motion of the atoms and the inelastic collisions of the electrons with each other and with the metal atoms."

Here, then, is one rather important particular in which the new theory is confessedly unsatisfactory.

As to the increase of conductivity which Bridgman's experiments have shown to accompany increased pressure on the metal, Sommerfeld is disposed to attribute this "not principally to the free path but to increase of n" [number of free electrons per cu. cm.]. A little further on he remarks, "our setting the number of electrons and the number of atoms equal to each other is only an indication of the order of magnitude; no exact equality is supposed to hold. Just as we regard n as dependent on pressure, we shall also in later formulas introduce a variability of n with temperature."

Since my own theory of metallic conduction, etc., has, I believe, sometimes been criticized as involving a considerable number of assumptions of a more or less special character, I am interested in these evidences that the new theory cannot get on without similar inventions. Variation of n with temperature has been one of my cardinal hypotheses.

On page 25 Sommerfeld says, "As to supra-conductivity our theory gives no solution; this seems to rest on an entirely different principle."

The Wiedemann-Franz Law.—Sommerfeld takes considerable satisfaction, naturally enough, in the success of his theory with the Wiedemann-Franz law. Thermal conductivity being indicated by κ, the Lorentz formula is

$$\frac{\kappa}{\sigma} = 2\left(\frac{k}{e}\right)^2 T.$$

The original Drude formula was

$$\frac{\kappa}{\sigma} = 3\left(\frac{k}{e}\right)^2 T.$$

The new theory gives

$$\frac{\kappa}{\sigma} = 3.3\left(\frac{k}{e}\right)^2 T,$$

which seems to be in somewhat better accord with the measurements of Jaeger and Diesselhorst than either of the other forms. A footnote, however, appears to intimate that newer and more exact measurements by Grüneisen and Goens confirm the Drude formula rather than the later one of Lorentz, or the latest one of Sommerfeld.

I venture to remark here, however, that each of these three formulas asserts too much in making the $(\kappa \div \sigma)$ ratio the same for all metals and strictly proportional to T. Experiment verifies neither of these propositions.

Thermionic Emission, the "Richardson Effect."—Even at the temperature of a glowing metal emitting electrons, Sommerfeld's criterion shows his "electron-gas" within the metal to be highly "degenerate." He gets as the emission formula

$$J = \frac{2\pi emG}{h^3} (kT)^2 \exp\left(-\frac{W_a - W_i}{kT}\right),$$

and remarks, "This equation agrees superficially with the Richardson formula as modified by Dushman, von Laue and Schottky, but with an important difference. The effective emission-work which occurs in the exponent of ϵ is not the "external" emission-work W_a, but the difference between "external" and "internal" emission-work, $W_a - W_i$. While W_a is the work to be done by the electron at emission (or shortly after the emission), W_i means an amount of work supplied to the electron during emission (or shortly before emission)."

"By what forces is the work W_i supplied? Evidently by the pressure of the electron-gas, which, according to the equation of condition of the electron-gas, exists within the interior of the metal and which with sufficient accuracy can be taken as the zero-point pressure." From his equations Sommerfeld finds W_i to be 5/3 times the mean kinetic energy of an electron within the metal.

I hope that what Sommerfeld has to say concerning the exponent of ϵ in the emission formula will have at least the good effect of leading others who are discussing emission to reconsider carefully their reasons for interpreting the numerator of this exponent in the ordinary Richardson empirical formula as the "work-function," pure and simple, of emission, what Sommerfeld calls the "external" work of emission of an electron. In my opinion the only justification for this interpretation is the assumption made by Richardson on page 33 of his "Emission of Electricity" (1916 edition) but afterward rejected, that we have within the metal a body of free electrons having the same mean energy of thermal agitation as the free electrons outside the metal. If this equality is not assumed to exist, it is difficult, I think, to attach any very definite meaning to

the numerator of the exponent of ϵ. The Richardson equation may in this case be regarded as merely an empirical one, and the relation of the quantity in question to certain other quantities of energy, such as the calorimetric work-function and the photo-electric work-function, should be sought by way of experiment.

Accordingly, I do not feel sure that Sommerfeld's $(W_a - W_i)$ is correct, even if his electron-gas has the properties he attributes to it, but, on the other hand, it might be difficult to show by theoretical reasoning that his conception of the quantity in question is wrong.

Neither Sommerfeld nor Houston, who writes a supplementary paper on emission, refers to the thermionic-emission work of Davisson[2] and Germer, from which these investigators have drawn the conclusion that the mean kinetic energy of the conduction-electrons within a metal is much *less* than the thermal energy of gas molecules at the temperature of the metal.[3] I was at first disposed to regard Sommerfeld's hypothesis of a very high kinetic energy of the conduction-electrons as incompatible with the experimental evidence here referred to. A closer examination of this evidence shows, however, that we cannot draw from it any inference regarding the kinetic energy of the conducting electrons without making some assumption which may not be justified concerning the ϕ of the Richardson formula. I do not, therefore, criticize the Sommerfeld theory at this point.

Electron Emission from Cold Metals.—This subject, Millikan and Eyring's "pulling" electrons out of metals, is dealt with by W. V. Houston on pages 33–37, supplementing the paper of Sommerfeld. According to Houston there is a general agreement between the indications of Sommerfeld's theory and the experimental results found by Millikan and Eyring, but he gives a number of reasons why a close comparison is not practicable at present. There is "a factor of more than 150 to be explained by irregularities and impurities of the surface."

Photo-Electric Emission of Electrons.—This subject is, I believe, not mentioned in the development of Sommerfeld's theory in the papers under discussion. We are left to imagine what Sommerfeld's conception of photo-electric emission is.

Volta Effect.—This subject is discussed on pages 38–42, by Carl Eckart, who begins thus: "It seemed at first as if the Volta effect presented difficulties[4] which by the theory developed in the first of these papers [Sommerfeld's part] could be overcome only by rather arbitrary assumptions. On closer examination, however, it appears that these difficulties vanish if one finally assumes that the Volta potential difference comes from an electrical field in the space that surrounds the metal." After reference to the notion of a "double layer" at the surface of the metal, to account for the work of mere emission, he continues: "But we shall

see that one must still add a surface charge which will produce a field at a greater distance from the metal surfaces."

The discovery that a Volta potential-difference between two metals implies an electric field in the space surrounding the metals is, of course, not a new one. Richardson was evidently familiar with this idea when he wrote his "Emission of Electricity" (see page 41 of the 1916 edition).

As one might expect from Sommerfeld's discussion of thermionic emispion, the part payed by ϕ_1 and ϕ_2 in Richardson's expression for the Volta Sotential difference is played by $(W_{a_1} - W_{i_1})$ and $(W_{a_2} - W_{i_2})$ in the sommerfeld-Eckart expression. When, however, Doctor Eckart comes to discuss the conditions that exist, or the forces that act, at the *junction* of the two metals he persuades himself that the difference between W_{a_1} and W_{a_2} must there vanish. How this can be, he does not profess to understand, but he argues at considerable length to show that it must be so. I have not space here to discuss his argument, which seems to me unconvincing.

We come now to the second part of Sommerfeld's own paper, pages 43–60.

Thomson Effect.—The formula derived by Lorentz for the Thomson effect coefficient, μ_{Cl}, is given, with the remark, "It [the value of μ indicated by this formula] is, compared with experiment, much too large, even larger than the abnormally great experimental values in bismuth, arsenic and antimony."

The new theory gives an expression for μ which, compared with the Lorentz formula, gives the ratio

$$\frac{\mu}{\mu_{Cl}} = \frac{4\pi^2}{9}\frac{mkT}{h^2}\left(\frac{4\pi G}{3m}\right)^{\frac{2}{3}}.$$

For silver at 300° absolute Sommerfeld finds this ratio to be about 0.01. He then remarks, "Thus we come to the right order of magnitude of the experimental values of Borelius and Gunneson for the noble metals copper, silver and gold. Moreover, the proportionality with T which our formula . . . demands is very well confirmed in the observations at high temperatures. Whether the departure of the observations from proportionality at lower temperatures agrees with our . . . equation . . . , we will not decide."

After some further reflections he sums up as follows: "We come therefore to the conclusion that the order of magnitude of the Thomson effect and its dependence on temperature are correctly given by the new statistics. The abnormal values in bismuth, arsenic and antimony naturally fall out."

In estimating the weight of this claim, which may prove to be a just one, it is well to note that the three confirming metals, copper, silver and gold, are very much alike in their physical properties, so that any theory that succeeds with one of the three is likely to succeed with all of them. More-

over, though arsenic may well be ruled out as not being strictly a metal, it appears to be somewhat arbitrary to rule out bismuth and antimony also, their only offense being, apparently, that their Thomson effects are "abnormally" large.

Peltier Effect.—After deriving his formula for Π, the Peltier effect coefficient, by the new theory, Sommerfeld comments thus: "A rough evaluation . . . shows that Π is equal to some hundreds of micro-volts, which corresponds very well with the available observations on normal metals. Noteworthy is the proportionality with T^2, while the classical formula . . . would require proportionality with T."

He gives no evidence to show that observation confirms this T^2 relation, unless an instance given under the next heading is to be so taken. Certainly, the Peltier effect between metals is not, in general, proportional to T^2.

Thermo-Electromotive Force.—For a thermo-electric couple with junctions at T' and T'', respectively, Sommerfeld gives these formulas for the net electromotive force, Θ:

Classical
$$\Theta = \frac{k}{e} \log \frac{n_2}{n_1} (T'' - T'), \tag{63a}$$

New
$$\Theta = \frac{\pi^2}{3} \cdot \frac{mk^2}{eh^2} (\lambda_1^2 - \lambda_2^2)(T''^2 - T'^2), \tag{63c}$$

the λ_1 and λ_2 being "de Broglie wave-lengths" in the new wave-mechanics.

It is well known that Θ is not, in general, strictly proportional to $(T'' - T')$, but certainly it is not, in general, proportional to $(T''^2 - T'^2)$.

The only test or illustration of this latter relation given by Sommerfeld is in the case of a sodium-silver couple, with $(T'' - T') = 1$ and $0.5\ (T'' + T') = 273$. In this case he gets from his equation (63*c*) the value $\Theta = -2$ micro-volts, approximately, adding, "The observed value is somewhat larger, but the order of magnitude is right."

Volta Series and Thermo-Electric Series.—"A difficult question," says Sommerfeld, "is that concerning the relation of the thermo-electric series to the Volta series." He does not accomplish very much in his brief discussion of it.

The paper which I presented to the Volta Congress at Como last September dealt expressly with the Volta effect and its relation, or lack of relation, with the Peltier effect. This subject is very closely related to the question here touched upon by Sommerfeld. I hope that when my Como paper appears in print some one will pay me the compliment of reading it as carefully and criticizing it as freely as I have done in the case of this paper of Sommerfeld.

Galvanometric and Thermo-Magnetic Effects.—Sommerfeld discusses

the Hall effect and the Nernst effect from the point of view of the new statistics but reaches no revolutionary results in this field.

[1] This interesting reflection upon Sommerfeld's discussion was first called to my attention by my colleague, Professor Bridgman.

[2] *Phys. Rev.*, **20**, 300–330 (1922).

[3] I have pointed out (*Proc. Nat. Acad. Sci.*, **13**, 43–46 (1927)) that the dual theory of electric conduction permits us to suppose the *free* to have the heat energy of gas molecules while the mean heat energy of all the conducting electrons may be much less.

[4] In the first draft of his theory Sommerfeld, using, I believe, purely kinetic considerations, deduced Volta potential-differences which appeared to be of the right order of magnitude but were, unfortunately, of the wrong sign. Doctor Eckart now amends the theory by introducing something like the Volta conception of a specific attraction of different metals for electricity.

ELECTRON "FREE PATH" AND SUPRA-CONDUCTIVITY IN METALS

BY EDWIN H. HALL

JEFFERSON PHYSICAL LABORATORY, HARVARD UNIVERSITY

Communicated April 14, 1928

Probably everyone who has thought much on the mechanism of metallic conduction has come to the conclusion that the so-called mean free path of the "electron-gas" particles within a metal must be taken as many times the "atomic distance"—distance from center to center of neighboring atoms. For example, Sommerfeld's recently proposed theory of conduction makes this path "some hundreds of time the atomic distance." A little reflection will show that a path of such length within a metal can bear no close analogy to the "mean free path" of the kinetic theory of gases—a straight-line movement from one collision to the next—unless the electrons are supposed to go freely through the atoms. If they do go through the atoms or if they merely zigzag their way among the atoms, what terminates their free path? Why shouldn't it extend indefinitely?

The classical formula for the free-electron conductivity of a metal is

$$\sigma = C \frac{e^2 n l}{\sqrt{T}}, \qquad (1)$$

where C is a constant, e is the electron charge, n is the number of free electrons per cu. cm. of the metal and l is the length of the so-called free path. I shall keep to this formula but give an unusual interpretation to l.

The length of l is significant as indicating, in connection with the velocity of to-and-fro motion of the electrons, how *long*, during each "path," an electron is subjected to the accelerating force of the potential gradient which maintains the current. It is not necessary that l shall be regarded as a straight-line path. It may for the purpose of *conduction*—maintenance of a net drift of the electrons in one direction—be a zigzag path with any number of deflections, provided the general direction of its course is maintained.

There are two ways in which we may think of such a path as ending. We may suppose the electron to have a collision that will *reverse* its general direction of motion, so that any increment of velocity it may have gained by reason of the applied potential gradient will, together with the rest of its velocity, be turned about, in which case the potential gradient will have to begin anew its impelling influence on the electron. This must, I suppose, be the ordinary conception of the limitation or termination of a free path, though most writers who speak of free paths do not undertake to tell just what they have in mind.

What I now propose is more definite. The free path terminates when the free electron is captured by a positive ion. This would seem to be the natural fate of a free electron. This conception of the termination of free path, though perhaps novel, seems to be inherently more probable, as well as more definite, than the alternative one. For it must be remembered that, if there are free electrons, there must be positive ions within a metal, and the natural tendency of these oppositely charged bodies is to pair with each other.

If we adopt this view of free path, some interesting consequences follow:

Specific Heat.—The length of a free path will, other things being equal, be inversely proportional to the number of ions per cu. cm., and, as these ions must be equally numerous with the free electrons, this means that nl is, other things being equal, a constant. In other words, the conductivity of a metal is, other things being equal, independent of the number of its free electrons per cu. cm. This disposes of the specific heat difficulty, of which so much has been made in criticism of the classical theory of free-electron conduction, for we can make the number of free conducting electrons as small as we please.

Dependence of Conductivity on Temperature.—Framers of conduction theories are usually pleased if their formulas make the conductivity inversely proportional to T. To be sure, any theory which proves this relation proves too much, for conductivity is not, as a rule, strictly proportional to T^{-1}; but something pointing in the general direction of such a rule is a desirable asset of any conduction theory.

My own theory, which, so far as the free-electron conductivity goes, makes use of equation (1), needs therefore to show how nl may be pro-

portional, pretty nearly, to $T^{-1/2}$. The theory assumes a fairly rapid increase of n with rise of temperature, but, as we have seen, this increase of n automatically shortens l to such an extent that, if this were the only cause reducing l, nl would be independent of temperature. I must, then, look for some other influence of increased temperature on the length of l, tending to reduce it.

It is possible that the increased thermal agitation of the positive ion with rise of temperature may have some effect upon the chance of its capturing an electron, but it is not plain what this effect should be, or even in which direction it would affect the length of l. Therefore, I prefer not to speculate upon this possibility. On the other hand, my theory, as developed several years ago, takes the work of ionization within a metal—the work, that is, of getting an electron out of an atom into the free state—as nearly proportional to the absolute temperature. This implies that the potential energy of a free electron with respect to a neighboring ion is pretty nearly proportional to the temperature, and evidently this is a condition tending to bring about an early capture of the electron, tending to reduce the length of l. Just how great the effect of this influence would be, it is not, perhaps, easy to say, but it seems not unreasonable to suppose that the effect of all the conditions suggested might make the conductivity pretty nearly proportional, inversely, to the temperature. It is doubtful whether any theory of conduction should be expected to indicate a closer relation between conductivity and temperature than this. If strict inverse proportionality were indicated, it would be necessary to invent some way of explaining why every individual metal, each in its own way, departs from this relation.

Supra-Conductivity.—Two of the hypotheses of my dual theory of conduction, which were adopted long ago for the temperature region between 0° and 100°C., are these:

$$n = zT^q \tag{2}$$

where n has the same meaning as in (1), while z and q are constants; and

$$\lambda' = \lambda'_c + skT \tag{3}$$

where λ' is the heat of ionization, with respect to a single electron, within the metal, while λ'_c, s and k are constants, the last being the gas-constant for a single molecule. For most metals the λ'_c term in (3) appears to be much smaller than the skT term.

By extrapolation toward the absolute zero, equation (2) indicates a very small number of free electrons at very low temperatures, but (3) indicates approach to a state in which, λ' becoming very small, the distinction between free electrons and "associated" electrons—that is, elec-

trons still in atoms but capable of going over to adjacent ions—becomes small. If we imagine a condition, at some low temperature, in which λ' becomes zero, the distinction between free electrons and associated electrons will be extinguished. In such a case the electrons, if once started in a given direction by an applied e.m.f., might go on indefinitely in that direction, there being nothing to stop them. That is, we should have supra-conductivity. I first made a suggestion to this effect,[1] though without the conception of free path that I have used in this paper, in 1926.

Thermal Conductivity at Low Temperature.—Since, according to my theory of the matter, thermal conductivity in metals depends mainly upon the dual character of electric conductivity and upon the heat of ionization, it should be especially small when, as in my conception of supra-conductivity, "free" electrons and "associated" electrons become indistinguishable, the heat of ionization becoming zero. It should not, however, be entirely lacking, as the elastic vibrations of the metal would have some heat-carrying power.

[1] See page 417 of *Phys. Rev.*, **28.**

DISPERSION OF LONG WAVE-LENGTH X-RAYS IN PLATINUM AND CALCITE

By Elmer Dershem

Department of Physics, University of California

Communicated April 2, 1928

The problem of indices of refraction and dispersion in the x-ray region was first brought to the attention of physicists by the observation that the apparent wave-length of x-ray lines measured in different orders did not agree and that the discrepancies were such as to be explained by the theory that the crystal used had an index of refraction slightly less than unity.[1] From these measurements the order of magnitude of the index could be calculated but the method was indirect and not capable of much accuracy. The results, however, showed the importance of finding the x-ray refractive indices of crystals in order to correct existing wave-length measurements for this effect.

A. H. Compton[2] attacked the problem in a more direct way by showing that polished surfaces would reflect x-rays if the glancing angle of incidence were sufficiently small and that a critical angle of total reflection could be measured and the index of refraction computed as in ordinary optics. He also showed that the indices measured in this way were in good agreement with an extension to these high frequencies of theories

of dispersion developed by Drude[3] and Lorentz[4] for the region of optical frequencies. This may be expressed in the following formulas:

$$1 - \mu = \delta$$

$$\delta = \frac{e^2N}{2\pi m} \sum_1^Z \frac{n_c}{\nu^2 - \nu_c^2}$$

in which μ is the index of refraction, e and m the charge and mass of an electron, N the number of atoms per cc., ν the frequency of the incident x-ray, ν_c a critical frequency of the atom, n_c the number of electrons in the atom having this critical frequency, and Z the total number of electrons in the atom or its atomic number.

If this formula applies then δ should have large positive values when the frequency of the incident x-ray is just slightly larger than one of the critical frequencies of the atom and large negative values for frequencies slightly smaller than a critical frequency. In other words, the phenomenon of anomalous dispersion should appear.

Hjalmar and Siegbahn[5] thought to have found evidence of anomalous dispersion from a consideration of the apparent differences of the same wave-length reflected from calcite and gypsum. von Nardroff,[6] Hatley[7] and Edwards[8] have sought to identify the x-ray absorption limit frequencies with the critical frequencies of the above formula.

Kronig[9] and Kallman and Mark[10] have shown that from a consideration of the type of absorption curve found near an x-ray absorption limit frequency, a dispersion formula may be developed which Doan[11] has put into the following form:

$$\delta = \frac{NZe^2}{2\pi m\nu^2} + \frac{Nc^4CZ^4}{8\pi^2\nu^4} ln \left(\frac{\nu_c^2 - \nu^2}{\nu_c^2}\right)^2$$

in which C is the constant of the fluorescent absorption coefficient and the other symbols have the meaning used above. The effect of the logarithmic term is to give lower values of δ on both sides of the absorption limit wave-length instead of lower values on the long wave-length side only and higher values on the short wave-length side as is the case with the Drude-Lorentz formula. For frequencies not near an absorption limit both formulas lead to the same values.

The experimental data on which these theories might be tested have previously been very meager and confined to a small number of isolated measurements of δ for a few favorable wave-lengths in the region of high frequency x-rays such as may be easily obtained from an ordinary x-ray tube in air. The formulas, however, indicated that larger values of δ should be obtained at longer wave-lengths, such as could be obtained only in a vacuum.

A large vacuum spectrograph was constructed to carry out this work. The evacuated bowl in which the apparatus was enclosed was two feet in inside diameter, being much larger than the usual vacuum x-ray spectrograph and thus permitting greater wave-length resolution. An especially designed water-cooled x-ray tube was connected to the bowl and the arrangement was such that a beam of x-rays entering the bowl from the tube was collimated by passing through two narrow slits 25 cm. apart. This beam then impinged upon a gypsum crystal at the center of the spectrograph and was partly reflected by it to a photographic plate tangent to a circle of 29 cm. radius having the gypsum crystal at the center. This reflected beam was nearly monochromatic and its wave-length could be determined from Bragg's law after the instrument had been calibrated by photographing the L spectrum of tungsten. A mirror surface of the substance to be investigated could then be introduced into this beam at a distance of 5 cm. from the crystal and 24 cm. from the plate, its angle of inclination varied and the maximum angle at which reflection occurred determined. The experimental values of δ were then calculated from the equation

$$1 - \mu = \delta = 2 \sin^2 \frac{\phi}{2}$$

ϕ being the maximum glancing angle at which reflection occurred. With this apparatus it was found possible to secure sufficient intensity even when the beam of general radiation was so narrow that variations from the mean wave-lengths were less than $^1/_5\%$. This ability to make use of general radiation removed one of the most serious limitations that had hampered all previous investigations in which only characteristic line radiation was used and permitted the measurement of δ using any wave-length within the range of crystal reflection. It is expected that a more complete description of this apparatus and the methods employed together with other data obtained will be published elsewhere, the purpose of this paper being to state briefly the results so far secured

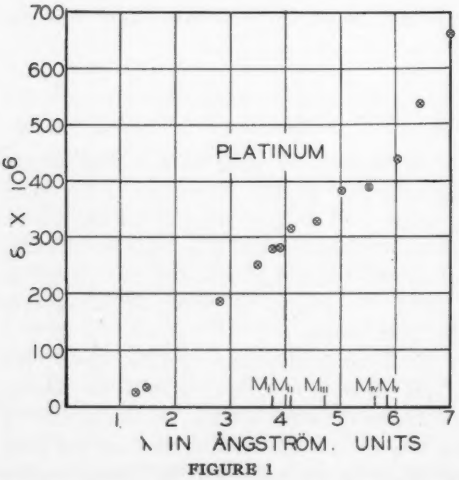

FIGURE 1

with platinum and calcite and the bearing of these results on the previously mentioned theories.

Platinum was first chosen as a reflector since smooth mirror surfaces which do not readily tarnish may be easily obtained of this metal by thermal sputtering upon plate glass. Platinum is also very dense and δ should, through a certain range at least, be approximately proportional to density. Also its M series absorption limit wave-lengths were within easy range of the instrument and it was thought that perhaps δ would undergo large changes as one passed to frequencies lower than those of these limits. The values of δ and the corresponding values of λ are shown in figure 1. It may be seen that while sufficient points have not been obtained to determine the fine structure of a dispersion curve the values of δ continue to increase with λ and to do this at an accelerated rate even after passing to frequencies lower than the characteristic frequencies of 28 of the 78 electrons of platinum. In other words, δ continues to increase and the index of refraction to become smaller although more than one-third of the electrons have characteristic frequencies greater than that of the incident x-ray. Carbon paper was used to protect the photographic plates from the light of the x-ray tube filament and as this appears to become nearly opaque for x-radiation at a wave-length of about 7 Ångstrom units the work was not carried, for the time being, to greater wave-lengths. However, from these results it would appear very probable that δ in the case of platinum continues to increase with increase in λ over a much greater range, at least as far as the beginning of the N series of absorption limits at about 17 Ångstrom units.

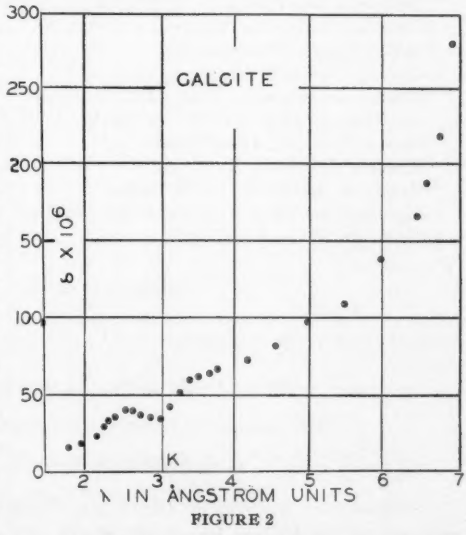

FIGURE 2

Calcite was next chosen because its use as a crystal grating made it desirable to study its dispersion curve in order that wave-lengths determined by its use might be corrected for refractive index effects. The K limit of calcium was also in a favorable region for study. Certain improvements in the technique of handling the spectrograph made it possible

to use narrower slits and thus a more nearly monochromatic source and at the same time work more rapidly. Figure 2 shows values of δ plotted against λ.

These results show very clearly the type of anomalous dispersion occurring near a K absorption limit. It is of the form predicted by the theories of Kronig and of Kallmann and Mark. It is doubtful that a quantitative agreement exists. For one thing the fluorescent absorption coefficient of calcium is not accurately known in this region. If the dispersion theory were quantitatively reliable perhaps the absorption coefficient could be derived from the dispersion curve. However, it must be noted that in the range of from six to seven Ångstrom units δ more than doubles in value. This is much more than can be explained by either theory.

[1] Stenstrom, also Hjalmar; see Siegbahn, *Spectroscopy of X-Rays*, p. 21, **1925.**
[2] Compton, *X-Rays and Electrons*, p. 205, **1926.**
[3] Drude, *Theory of Optics*, p. 382.
[4] Lorentz, *Theory of Electrons*, 2d ed., p. 150.
[5] Hjalmar and Siegbahn, *Nature*, **115**, 85 (1925).
[6] von Nardroff, *Phys. Rev.*, **24**, 143 (1924).
[7] Hatley, *Phys. Rev.*, **24**, 486 (1924).
[8] Edwards, *Phys. Rev.*, **30**, 91 (1927).
[9] Kronig, *J. Optical Soc.*, **12**, 547 (1926).
[10] Kallmann and Mark, *Ann. Physik*, **82**, 585 (1927).
[11] Doan, *Phil. Mag.*, **4**, 100 (1927).

MOBILITIES OF GASEOUS IONS IN SO_2 AND SO_2-H_2 MIXTURES

BY L. DUSAULT AND LEONARD B. LOEB

DEPARTMENT OF PHYSICS, UNIVERSITY OF CALIFORNIA

Communicated April 9, 1928

Introduction.—In order to throw possible light on the nature of gaseous ions one of the writers began the study of mobilities of ions in mixtures of various gases of diverse chemical nature.[1,2,3] This study constitutes a continuation of the work extending it to SO_2-H_2 mixtures. Despite the fact that mobilities have already been measured in pure SO_2 the results obtained in the pure gas alone merit some discussion.[4,5] While theoretically an easy gas to work with and assumedly easy to prepare in pure form, it was found to be the most difficult gas yet encountered. The results are therefore not entirely satisfactory, but inasmuch as the work must be temporarily discontinued, it seemed best to report the results obtained to date.

Effect of Fine Gauzes in Mobility Measurement.—Before undertaking measurements the writers briefly investigated the question of the nature

of the gauze to be used in the Franck modification of the Rutherford Alternating Current method. One of the writers[6] had previously investigated the effect on mobilities due to the interpenetration of the auxiliary and alternating fields through perforated sheet brass gauzes. In the past, such gauzes had been used by the writers in their mixture work, but these necessitated corrections for the ratios of the auxiliary and alternating fields used, which were bothersome. According to recent articles of Schilling[7] and Herbert Mayer,[8] fine mesh gauzes having fine wires do not produce the changes in mobility with changing values of the above ratio observed by the writer. A number of gauzes of this type were prepared,

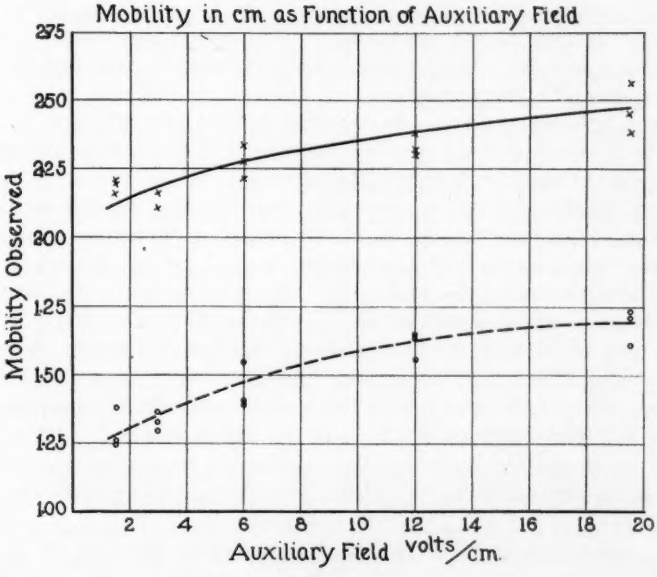

FIGURE 1

and one having meshes 0.33 mm. on a side and wires about 0.1 mm. in diameter was chosen. Such gauzes can be prepared by taking 80-mesh phosphor bronze screening and etching out the wire with HNO_3. The gauzes so prepared are as fine as any gauzes previously used. By proper mounting they can be made plane and have a very high transmitting power for ions, at the same time having small meshes.

The mobilities were measured in H_2 and in pure dry air in the carefully cleaned ionization chamber used before,[11] employing different values of the auxiliary field. Above auxiliary fields of 12 volts per cm. the positive and negative mobilities in air (and also in H_2) were high and

relatively constant up to 20 volts per cm. and presumably higher. Below 12 volts per cm., auxiliary field strength, the mobilities dropped more rapidly. The alternating fields lay in the neighborhood of 30 volts for the negative ions and 40 volts for the positive ions. The values obtained are shown graphically in the curves of figure 1 for air, several different specimens of air being used causing the spread of results. Analogous results were obtained in H_2. The shape of the curves indicates a more rapid rise of the mobilities of both ions as the field strengths are first increased than later. It is probable that beyond 20 volts auxiliary field strength to 80 volts the change would be relatively slight and possibly masked by the smaller irregularities introduced by the incompletely controllable purity of the air used.

Thus, contrary to the assertions of the authors mentioned,[7,8] such fine gauzes also show the variable mobility previously described by one of the writers.[6] The mobility curves obtained at different fields show the same behavior as previously described and the only difference is that the effects occur at lower auxiliary field and, therefore, higher ratios of alternating to auxiliary field with the finer mesh gauzes.

Thus Schilling's apparent constancy was due to a limited range of variation of auxiliary field on the flatter portion of the curve. His high absolute values for positive and negative ions in air are then to be ascribed to the high auxiliary fields used. This is borne out by the fact that the writers observed mobilities just as high as those of Schilling at the fields used by him, on air that he would designate as impure. This is precisely what had been anticipated by one of the writers elsewhere.[12] The conclusion to be drawn from this is that the method is incapable of giving absolute mobilities of ions, and the values obtained must be reduced to an absolute scale through comparison with a more accurate method, or else by calibration with a gas whose absolute mobilities are known.

The measurements to be described were carried out with the gauze discussed above using an auxiliary field of 12 volts/cm. By comparison with air the true values on the new absolute scale can then be obtained by multiplying the observed values by 0.96, and the values on the old recognized scale by multiplying by 0.79.

Preparation of the SO_2.—The SO_2 was obtained from two sources. In the first case, a sample of the best commercial Na_2SO_3 to be obtained from the Chemistry Department was placed in an evacuated flask and C. P. Concentrated H_2SO_4 was dropped onto it. The SO_2 generated passed through a trap cooled to $-70°$C. in frozen alcohol, through two previously exhausted tubes of P_2O_5 each a meter long, through another double cooling trap at $-70°$C. and into a reservoir where it was frozen out with liquid air. In the second case, owing to the peculiar results

first obtained, the SO_2, on the advice of the Chemistry Department, was taken from a commercial bomb containing liquid SO_2, the first half of which had already been boiled off. After freezing, the bulb where the SO_2 collected was pumped to the best vacuum obtainable with a mercury vapor pump. In the case of the first sample this was fractionated in use, while the second lot was purified by boiling off half of the 50 cm.[3] condensed out in the trap in filling. As far as the writers could foresee, this gas should have been quite pure. There was a trace of Hg vapor in the apparatus due to the mercury gauges. The only other source of possible contamination was the rubber stopcock grease (Lubriseal) used on the stopcocks, which could not be prevented in this system from coming in contact with the SO_2 to a slight extent. After considerable use, on evacuation and washing once with dry H_2, the ionization chamber on opening had no odor of SO_2 and only a very faint suggestion of an aromatic substance. Occasionally in measurement the last traces of SO_2 were removed with NH_3 gas, but this did not notably change the control mobilities in H_2 if care was taken to pump out the NH_3.

Results in SO_2 Alone.—The measurements in SO_2 gave the following values in cm./sec. per volt/cm. at atmospheric pressure under the conditions used:

	POSITIVE IONS	NEGATIVE IONS	WHEN MEASURED
1	0.346	0.332	First measurement
2	0.347	0.330	Fifth measurement
3	0.352	0.337	End of the first lot of SO_2
4	0.354	0.307	Beginning of second lot of SO_2
Average	0.350	0.329	
5	0.504	0.462	Central section of first lot of SO_2

The values numbered 1 and 2 were taken near the beginning, 3 was taken near the end of the first charge of SO_2 but after number 5 below had been taken, and 4 was found on the new charge of SO_2 from the commercial bomb. The fifth measurement gave rather surprising values of 0.504 cm./sec. for positive ions and 0.462 cm./sec. for negative ions. These high values were not accidental and had to do with a particular state of purity attained in the center of the first charge of SO_2, which also characterized itself in other ways.

Absolute Values.—SO_2 in samples 1, 2, 3 and 4 corrected for the auxiliary field, gave mobilities of 0.276 and 0.260 cm./sec. on the old accepted scale and 0.336 and 0.316 cm./sec. on the new absolute scale, for positive and negative ions, respectively. The values on the probably pure sample indicated under 5, reduced to the old accepted scale, are 0.398 and 0.365 cm./sec. and 0.484 and 0.443 cm./sec. on the new absolute scale for positive and negative ions, respectively. The only other values in SO_2 were those of Wellisch[4] and Yen.[5] Both made measurements at low pressure 60–

200 mm., using the same method as the one above. Their results have little significance as absolute values, as the fields and gauze were not defined. Their values are 0.415 and 0.414 for Wellisch and 0.412 and 0.414 cm./sec. per vol. cm. for Yen for positive and negative ions, respectively. These are distinctly higher than the values of the writers in the cases 1, 2, 3 and 4 above and lower than the values in case 5 above. Yen[5] definitely asserted that the negative mobility was greater than the positive. This is in direct contradiction to the results above. In only one case in some forty measurements was the negative mobility observed greater, the results in all cases being such as to leave no doubt.

Results at Lower Pressures.—The present measurements were carried down to lower pressures with the results shown below. The table contains the mobilities as observed directly reduced to 760 mm. with no correction for the auxiliary field.

PRESSURE IN MM.	POSITIVE IONS	NEGATIVE IONS	NUMBER OF MEASUREMENTS
10	0.399	0.385	3
21	0.377	0.364	1
51	0.372	0.398	1
149	0.372	0.347	2
316	0.372	0.339	2
760	0.350	0.329	4

It is seen that there appears to be a real increase in the reduced mobility as the pressure decreases. This appears to be in agreement with a prediction of A. P. Alexeievsky.[9] It explains the higher values of the mobilities observed by Wellisch and Yen compared to the values observed by the writers as the former worked at low pressures. This, if correct, is an interesting deviation from the inverse pressure law. The measurements in mixtures seem to indicate that the probable explanation of this increase is no real deviation from the inverse pressure law, as Alexeievsky's theory would have it, but is due to the fact that the smaller samples of SO_2 distilled over at lower temperatures have a higher average purity corresponding to the exceptional sample measured in measurement 5 of the set at atmospheric pressure. In all but case number 5 at atmospheric pressure the results indicating the purer specimens of SO_2 came for small amounts of SO_2, distilled over, say up to 30 mm. pressure in the chamber. Although some forty or more measurements were made in this study of SO_2, no control over this variable purity could be obtained, except that the pure samples occurred in the central fractions only and for distillates obtained at the lower temperatures. This type of difficulty is unique for, so far in their experience, the writers experienced no such difficulty which could not be explained or controlled.[3] It is possible as stated that the SO_2 reacted with the rubber stopcock grease to give a volatile poison. The possibility of SO_2 as an active impurity was not considered although as

later learned this gas is often present in appreciable amounts according to chemists. According to Zohn similar difficulties were encountered in his work on the dielectric constant of SO_2. It is hoped at a later date to revert to this problem and discover the nature of the difficulty.

Free Electrons.—The negative carriers in SO_2 down to 1 cm. were always negative ions of mobility less than the positive ions. To test this still further pure hydrogen showing free electrons at 760 mm. was mixed with 1 cm. of SO_2. The mixture was then pumped down to 3.8 cm. and tested for free electrons. None were observed. Thus a partial pressure of 0.05 cm. of SO_2 in 3.8 cm. of H_2 removes all free electrons. This leads one to conclude that electrons attach to SO_2 to form negative ions nearly as readily as they do to Cl_2.

Results in SO_2-H_2 Mixtures.—The H_2 used in these measurements was similar to that used in previous work. It could be purified to give free electrons at 760 mm. pressure at will by heating a copper tube through

% SO_2	POSITIVE IONS	NEGATIVE IONS	NUMBER OF MEASUREMENTS
0.00	6.55	8.90	5
0.01	12.2	6.10	1
0.079	14.4	6.58	4+ and 5−
0.25	11.9 4.94	7.03	2 (1 high)
0.49	16.5 5.64	6.58	5 (1 high)
1.15	20.3 4.80	5.50	3 (1 high)
2.62	22.7 3.76	3.99	5 (2 high)
4.92	2.99	3.13	3
10.2	2.09	2.02	1
25.3	1.12	1.06	1
35.0	0.866	0.802	1
50.6	0.602	0.586	1
75.7	0.434	0.378	1
92.0	0.361	0.355	1
100.0	0.350 0.505	0.329 0.462	5 (1 high)

which it passed. It was used in this work without the heating to get negative ions to work with.[3] The mobilities obtained with an auxiliary field of 12 volts per cm. were 6.55 cm./sec. for positive ions and 8.73 cm./sec. for negative ions. The values of the mobilities in SO_2 used in these computations were 0.350 cm./sec. and 0.504 cm./sec. for the positive ions and 0.329 for the negative ions. These were the mobilities directly observed and were comparable to those in the mixtures. The reason for the use of the faster mobility of the positive ion in computation will be seen when the results are given. The values for the mobilities are given

below. These are average values, the number of measurements used being indicated. They are uncorrected for the auxiliary field, being, however, reduced in each case to 760 mm. pressure. The values averaged lay on the whole fairly close together and no such extremes as shown above were averaged.

To study these better they are plotted in the curves of figures 2 and 3, the reciprocals of the mobility being plotted as ordinates against the percentage composition of the mixture as abscissae. The light dashed curves and circled points represent the positive ions and the light full lines and

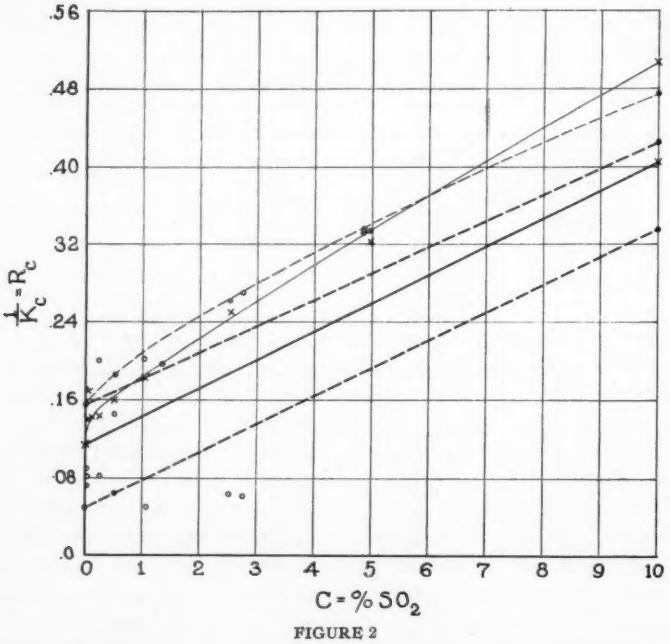

FIGURE 2

crosses the negative ions. Figure 2 goes from 0 to 10%, while figure 3 extends from 0 to 100% of SO_2. There are for figure 2 two sets of positive ion data due to the peculiar effect of the purer samples of SO_2 on the mobility of the positive ion in SO_2-H_2 mixtures in small amounts. The straight lines represent the computed curves from the mobilities in pure gases to be expected if Blanc's[10] law of mixtures holds. The two heavy dashed straight lines in figure 2 for positive ions apply to the computed curves for positive ions for the pure samples and for the normal samples.

If one represents the reciprocal of the mobility in a gas a by $\frac{1}{K_a} = R_a$,

and that for a gas b as $\frac{1}{K_b} = R_b$, and if in a binary mixture there is a mol fraction c of the gas a and $1 - c$ of the gas b present, the reciprocal of the mobility in the mixture $R_c = \frac{1}{K_c}$ will be obtained by merely adding the quantities R_a and R_b multiplied by the mol fractions of the respective gases present. Thus $R_c = cR_a + (1 - c)R_b$, which represents the

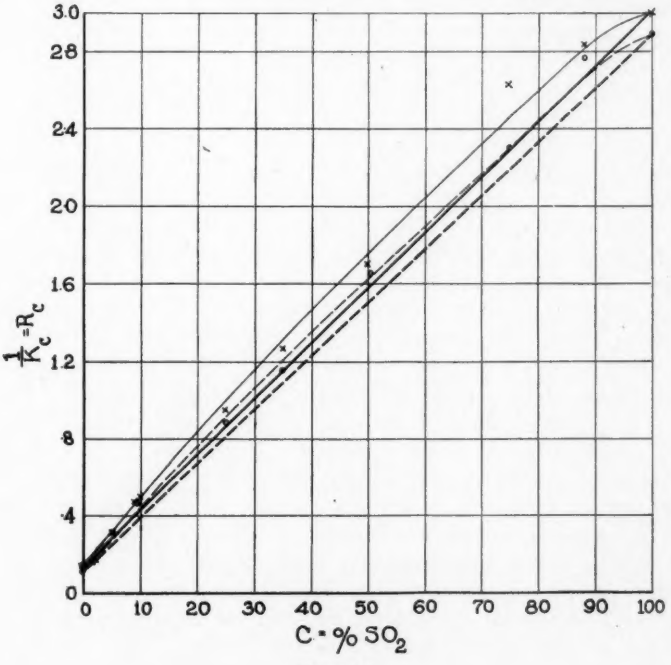

FIGURE 3

straight lines plotted using the observed end values. If the mobilities be substituted the law becomes

$$\frac{1}{K_c} = \frac{c}{K_a} + \frac{1 - c}{K_b}, \text{ whence}$$

$K_c = \frac{K_a K_b}{(1 - c) K_a + cK_b}$, which is the form of Blanc's law used in discussing previous work.[2,3]

Looking at the curves of figure 3, it is seen that these depart from the straight line law in a sense indicating that either a labile clustering or concentration effect is taking place. These deviations are quite like those

observed in NH_3 and represent a possible dielectric effect. In looking at figure 2, it is seen that SO_2 behaves like NH_3 in its action on the negative ion and also for the positive ions in the less pure samples. In the samples which have so far been specified as the purer samples, the positive ion has its mobility *increased to high values by traces of* SO_2 and even by amounts of SO_2 as great as 2.5%. The only possible interpretation of this is that in certain samples of the SO_2 free from some impurity SO_2 replaces the heavier molecule or molecules associated with the positive ion in H_2, making a more mobile ion. The highest mobilities observed with certainty for positive ions in these low percentage mixtures of SO_2 (they were not observed at higher percentages) were 22 cm./sec. per volt/cm. *This is higher than any negative ion mobility ever observed in* H_2. Negative ion mobilities as high as 13 have been observed in high auxiliary fields in fairly pure dry H_2. The effect of *NH_3 in increasing the positive mobility in H_2* was, on one occasion in these measurements, found to give a mobility as high as 14. Thus SO_2 is unique in the magnitude of its effect on the positive ion mobility in H_2. The interpretation of this is quite simple. It is certain that the negative ion in H_2 is formed by attachment of the electron to some molecule of impurity. If it were a stopcock grease or other bulky molecule present, this would even, in the absence of further clustering, give rise to a slower ion. Again in H_2 the ionization of H_2 by the alpha particles will lead to the formation of either H_2^+ or H^+. It is very doubtful if the positive charge will remain as long as 0.002 second attached to the hydrogen. The traces (from a chemical viewpoint) of impurities will quickly acquire the charge because of the lower ionizing potential. It may also be that the H^+ or H_2^+ will rapidly pick up some heavier molecule of impurity. In any case the positive ion will also be bulky, and apparently more so than the negative, for its mobility is less. Addition of pure SO_2 introduces molecules which have a stronger affinity for the positive ion than the impurity present before and the mobility is that due to a lighter complex. That the mobility should exceed that of the negative ion is not strange, as the latter is also formed about an impurity. Whether the charges are attached to single molecules of impurity or SO_2 in these cases it is impossible to say. The high values observed in SO_2 compared to the negative mobilities make it less likely that one is dealing with single molecular carriers than in the case of NH_3, where the mobility about equalled the negative. The sensitivity of SO_2 to some unknown impurity, however, shows that there are substances present that are even more effective than SO_2. This was also found before in the case of ether and NH_3, the ether being more active on the positive ions than NH_3.

The results, therefore, add one more striking bit of evidence indicating that besides general dielectric lowering effects, concentration effects or

statistical loose clustering, there are very much more stable complexes formed by the action of charges of one sign or the other on specific molecules. How large these are is still a question of doubt, but that such complexes must form cannot be questioned. Such combinations appear more or less like the complex ions encountered in chemistry, e.g., NH_4^+, $Cu(CN)_2^-$, etc., than like clusters in their specific character.

Summary.—The work is an extension of the investigations of ion mobilities in mixed gases to SO_2-H_2 mixtures. It is shown that the fine-meshed gauzes such as used by Schilling and H. Mayer show the effect of varying the auxiliary field on the mobility just as other gauzes do, differing only in the point where the effects become appreciable. This confirms previous work of one of the writers and leads one to suspect that Schilling's high mobilities are due rather to the high auxiliary fields used than to the exceptional purity. Difficulty was encountered in obtaining a purity as shown by these measurements which was satisfactory. Two sets of values of mobilities were observed at atmospheric pressure in SO_2. The absolute values on the new absolute scale for the supposedly less pure sample were 0.34 cm./sec. and 0.32 cm./sec., respectively, for positive and negative ions. For the other sample the values 0.48 cm./sec. and 0.44 cm./sec. on this same scale were observed. The mobilities reduced to 760 mm. appeared to increase with decreasing pressure a slight amount. The effect is probably ascribable to the higher purity obtained at lower pressures. Negative electrons could not be detected down to 0.05 cm. pressure of SO_2 in 3.8 cm. of pure H_2. The negative mobility is unquestionably less than the positive and appears so invariably in these measurements. The effect of SO_2 at higher concentrations in SO_2-H_2 mixtures is similar to the effect of NH_3 in NH_3-H_2 mixtures. At low concentrations the supposedly purer samples of SO_2 gave positive mobilities in H_2 of 22 cm./sec. per volt/cm. which exceed materially the negative mobilities observed. This must be ascribed to the formation of a relatively simple and mobile positive ion in SO_2 mixed with H_2. The negative ion as in NH_3 is unaffected except to have its mobility slightly reduced.

[1] Loeb and Ashley, *Proc. Nat. Acad. Sci.*, **10**, 351 (1924).

[2] Loeb, L. B., *Ibid.*, **12**, 35 and 42 (1926); **12**, 617, 677 (1926).

[3] Loeb and DuSault, *Ibid.*, **13**, 510 (1927); **14**, 193 (1928).

[4] Wellisch, E. M., *Phil. Trans. Roy. Soc. London*, **A209**, 249 (1909); *Phil. Mag.*, **34**, 59 (1917).

[5] Yen, K. L., *Proc. Nat. Acad. Sci.*, **4**, 106 (1918).

[6] Loeb, L. B., *J. Franklin Inst.*, **196**, 537 and 771 (1923).

[7] Schilling, H., *Ann. Phys.*, **83**, 23 (1927).

[8] Mayer, Herbert, *Phys. Zeits.*, **28**, 637 (1927).

[9] Alexeievsky, A. P., *Phys. Rev.*, **27**, 811 (1926).

[10] Blanc, A., *J. Physique*, **7**, 825 (1908).

[11] Loeb and Cravath, *J. Opt. Soc. Amer.*, **16**, 191 (1928).

[12] Loeb, L. B., *Ann. Phys.*, **84**, 691 (1927).

IS PITUITARY SECRETION CONCERNED IN THE INHERITANCE OF BODY-SIZE?

By R. Cumming Robb*

Bussey Institution, Harvard University

Communicated April 5, 1928

A causal relationship between the activity of the hypophysis cerebri and the growth of the body is indicated by the clinical association of acromegaly and gigantism (in man) with a pathological modification of this organ, first noted by Marie in 1886 (Marie and Marinesco, '91). This correlation has been sustained by the experimental production of gigantism in the salamander, *Amblystoma tigrinum*, by Uhlenhuth ('21), and in the rat by Evans and Long ('22), through the agency of anterior lobe substances. These data have given rise to the theory that normal differences in stature and body-build are referable to differences in the supply of pituitary, or thyroid, or some other endocrine substance available during the growth period. And on this premise it has been suggested that these organs represent the mechanism whereby mendelian determiners for body size make manifest their characteristic effects. Thus, in particular, it was supposed that the greater and more prolonged growth activity of some animals might be due to a proportionately greater endowment of pituitary "growth-promoting" substance. In order to test these hypotheses an examination has been made of the pituitary bodies in distinctly "giant" and "dwarf" races.[1]

Material and Methods.—In an investigation concerning the inheritance of body-size in rabbits Professor W. E. Castle has employed two races differing greatly in the velocity of growth and in ultimate body-weight. Whereas the "Polish" rarely exceed fifteen hundred grams, the "Flemish Giant" may attain a weight of more than six thousand grams. It is with great satisfaction that I express my indebtedness to Professor Castle for the opportunity to make, as an extension of his general problem, a survey of the endocrine glands in these two races to discover if possible the mechanism concerned in the determination of body-size.

For this purpose one hundred and fifty-two animals have been used, sixty-seven Flemish Giants, forty-seven Polish, thirteen Flemish-Polish hybrids and twenty-five rabbits of other genetic stocks,[2] which at maturity weigh between two and three thousand grams. These animals were all males and of selected ages from birth to senescence. This age distribution was such as to indicate for the entire post-natal life the gross relations of the endocrine glands in giant and dwarf male rabbits.

The animals were bled to death after a brief administration of ether; the roof and walls of the cranium were removed with bone-forceps, and the

brain slightly elevated in order that the infundibular stalk might be transsected. The bony wall of the sella turcica must be broken away and then the pituitary body may be dissected free from the enveloping membrane, placed in a small stoppered glass bottle and weighed to tenths of a milligram.

Owing to the storage of material in the intestinal tract the true weight of a rabbit (like its true rate of basal metabolism) cannot be obtained, at least during life. As a more accurate basis of comparison I have used the cleaned body weight. This value is obtained by subtracting from the live body weight the weight of stomach and intestine. The error involved by thus neglecting the tissues of the intestinal tract in computing body weight is constant and approximates 4.5%. The estimated error of the observations may be greater than this. For example, varia-

TABLE 1

MEAN WEIGHT IN MILLIGRAMS OF THE PITUITARY IN RABBITS OF BODY-WEIGHTS INDICATED, BELONGING TO THE FLEMISH GIANT RACE, THE POLISH RACE AND TO FLEMISH-POLISH F_1 HYBRIDS

CLEANED BODY-WEIGHT IN GRAMS	MEAN WEIGHT OF PITUITARY IN MG. FLEMISH GIANT NO.	MEAN	HYBRID NO.	MEAN	POLISH NO.	MEAN	THEORETICAL WEIGHT IN MG.
100, i.e. (50–149)	11	3.46			4	3.55	3.50±0.4
300 (150–449)	5	6.84			6	7.48	6.8±0.9
600 (450–749)	10	11.38			7	10.4	10.3±1.5
900 (750–1049)	5	12.9			16	13.3	13.0±2.0
1200 (1050–1349)	2	16.9			13	15.1	15.5±2.3
1500 (1350–1649)	3	21.6	1	19.2			17.7±2.7
2000 (1650–2349)	3	22.6	10	21.98			21.0±3.0
2500 (2350–2649)	3	22.63	2	24.1			24.0±3.6
3000 (2650–3349)	11	26.5					26.7±4.0
3500 (3350–3649)	2	27.4					29.0±4.6
4000 (3750–4350)	3	28.3					31.5±5.0
Totals	58		13		46		

tions in the level of sectioning of the infundibular stalk are inevitable and may cause a variation of five per cent in the pituitary weight. Incomplete removal of the enveloping membrane may cause an error of six per cent, weighing errors one per cent, the total being less than twelve per cent.

With regard to the technique involved in these dissections, it is a pleasure to acknowledge the kindly assistance rendered by Dr. F. S. Hammett, Director of Research at the Lankenau Hospital in Philadelphia.

Observations.—The data obtained for the weight of the pituitary body in giant, dwarf and hybrid rabbits are presented as individual records in figure 1. Pituitary weight is plotted against the "cleaned body-weight," a logarithmic scale being used for both. It is apparent that the triangles, which represent the dwarfs, are as a group not separable from the squares, which represent the giants. These data are also pre-

sented in table 1 as mean values, together with the theoretical means for all the data. In table 2 the weight of the pituitary is given as a percentage of the total body-weight. The conclusion is unavoidable that whatever may be the functional relationship of the hypophysis to body-growth the pituitary body as a whole does not show any characteristic differences of mass that may be correlated with the observed differences *in growth rate.*

TABLE 2

RELATION OF PITUITARY WEIGHT TO TOTAL BODY-WEIGHT

CLEANED BODY-WEIGHT IN GRAMS	PITUITARY AS PERCENTAGE OF BODY-WEIGHT GIANT	HYBRID	POLISH
100	0.0035		0.0036
300	0.0022		0.0025
600	0.0019		0.0017
900	0.0014		0.0015
1200	0.0014		0.0013
1500	0.0014	0.0013	
2000	0.0011	0.0011	
2500	0.0009	0.0010	
3000	0.0009		
3500	0.0008		
4000	0.0007		

Despite the evident decrease in the ratio of pituitary mass to body mass during the growth of the individual, there exists a very definite relationship between these values. *The rate of change in one is proportional to the rate of change in the other.* If the logarithm of the pituitary weight be plotted against the logarithm of body-weight, the points form a straight line. This relationship is identical in giant, dwarf, hybrid and nondescript male rabbits. Furthermore, the coefficient of variation is constant, the points forming a ribbon of uniform width. This suggests that the same group of growth "pace-makers" are concerned throughout postnatal life. (See Fig. 1.) The existence of the identical relationship between pituitary-weight and body-weight in both giant and dwarf animals indicates, not that the giant organism is a uniform enlargement of the dwarf, but that

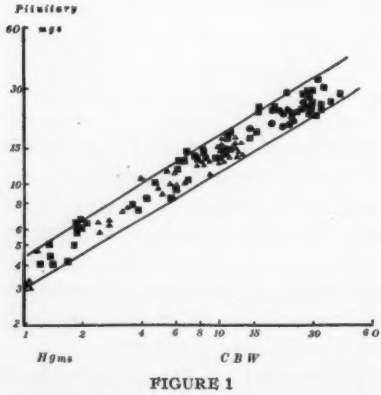

FIGURE 1

Weight of the pituitary body of the rabbit, in milligrams, against the "cleaned body-weight" in hectograms (logarithmic plotting both ways). Squares represent the Flemish Giants, triangles the Polish "dwarfs" and circles the F_1 hybrids, males only.

in it is found a further differentiation of certain tendencies already observed in the smaller animals.

Discussion.—The marked correlation between the rate of growth of the hypophysis and that of the total body-mass may be interpreted either as a "chance" association of no significance, or as manifesting a causal relationship. The former alternative, in the light of evidence cited above (and the known rôle of the pituitary substances in growth), does not seem probable.

If a significant relationship be admitted, these data indicate either the regulation of one by the growth of the other, or else that both these growth rates are themselves determined by another more fundamental process. The experiments of Uhlenhuth ('21) and of Evans and Long ('22) indicate that *within certain limits* the body will grow in proportion to the supply of pituitary anterior lobe substance available. These data apparently confirm such a deduction. But perhaps these data prove too much. The values that are here discussed are not a direct measure of growth-promoting substance. The anterior lobe represents only about seventy per cent of the pituitary mass (Björkman, '14) and to even this portion several different functions have been ascribed (Smith, '27; Flower, et al., '23; Cushing and Davidoff, '27). Apparently tissues other than the "growth hormone" producing acidophil may maintain a precise relationship to body-mass. Several other bodies in the rabbit—the thyroid, the adrenals and, prior to maturity, the thymus—are logarithmically proportional to body-mass; this I shall presently report in more detail. A similar relationship is found upon analysis of published data on the hypophysis, thyroid, adrenals, liver, skeleton and eyeballs of the rat (Donaldson, '24, tables 126, 144, 145, 148. For the convenience of the reader these data are presented

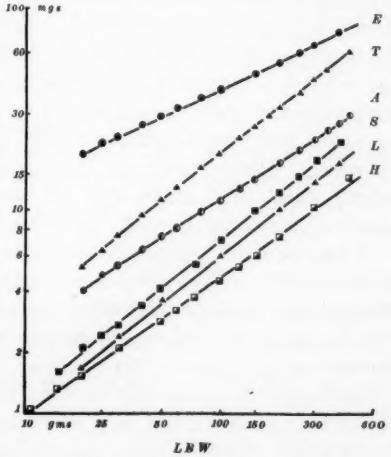

FIGURE 2

Weight of organs and the fresh skeleton in the rat, data from Donaldson ('24), plotted against the live body-weight (horizontal values), to show the post-natal rectilinear relationship. Logarithmic spacing both ways.
E, weight of eyeballs, in mgs. (× 2).
T, weight of thyroid in mgs.
A, weight of adrenal bodies in mgs. (× 1/2).
S, weight of fresh skeleton in grams.
L, weight of liver, also in grams.
H, weight of hypophysis, in mgs.

in Fig. 2). The growth of all these tissues during a more or less extended period of post-natal life is logarithmically proportional to the rate of body-growth. Are we to have a multiplicity of "causes" regulating body-growth?

The alternative seems the more logical interpretation, that the correlation between the growth of the organism as a whole and that of specific portions, such as the hypophysis, liver, skeleton, etc., is due to a common cause. It is true that an excess or deficiency of "pituitary growth-promoting substance" in those organisms normally possessing a pituitary does modify body-growth *within certain limits*. And concerning these limits it is hoped that experimental evidence may be obtained quite soon. But in view of the evidence presented in this paper it seems exceedingly probable that the basic factors which regulate the growth of the body are likewise concerned in limiting the growth of the pituitary. Moreover, the phenomenon of genetic growth determination is observed in plants and lower animals in which no endocrine glands have been identified. Hence, until specific evidence to the contrary can be presented, these data affirm that the interpretation of size-inheritance in terms of this endocrine body is not justified.

It may be further concluded that the logarithmic uniformity of growth manifested by various parts of the body in relation to the body as a whole indicate that their growth-rates are collectively determined by a single "master reaction" of which the magnitude is established by genetic determiners in a manner yet unknown.

Summary.—A comparison has been made of the weights of the pituitary bodies in giant (Flemish) and dwarf (Polish) male rabbits and in their F_1 hybrids. There is no characteristic difference in the mass of the pituitary that may be correlated with the observed differences in growth rate.

With increase in body weight there occurs a progressive decrease in relative pituitary weight. Accordingly, there is present in the full-grown dwarf relatively twice the amount of pituitary substance observed in the adult Flemish Giant.

Throughout post-natal life the weight of the pituitary maintains in all cases an identical rectilinear logarithmic relationship to body-weight. A similar relationship is characteristic of the growth of the thyroid, liver, skeleton, eyeballs, suprarenals and pituitary of the rat. This relationship between the rate of growth of parts and the rate of growth of the whole suggests their common regulation by a general master reaction of growth.

Björkman, H. (1915–16). "Bidrag till hypofysens aldersanatomie hors kainen," *Upsala Lakareforenings Forhandlinger, N. F.*, **21**, 49–108. (Cited by A. T. Rasmussen, '21, *Endocrinology*, **5**, 33–66.)

Cushing, H., and L. M. Davidoff ('27). "The Pathological Findings in Four

Autopsied Cases of Acromegaly, with a Discussion of Their Significance," *Monographs of the Rockefeller Inst.*, No. **22**, 131 pp.

Donaldson, H. H. ('24). *The Rat*, data and reference tables, 2nd ed., 469 pp.

Evans, H. M., and J. A. Long ('22). "Characteristic Effects upon Growth, Estrous and Ovulation Induced by the Intraperitoneal Administration of Fresh Anterior Hypophyseal Substance," *Proc. Nat. Acad. Sci.*, **8**, 38.

Flower, C. F., C. E. Forkner, W. E. Kellum, A. T. Walker, P. E. Smith and H. M. Evans ('23). "Separation of the Principle in the Anterior Hypophysis Affecting Ovulation from That Controlling General Body Growth," *Anat. Rec.*, **25**, 107.

Marie, P., and G. Marinesco (1891). "Sur l'anatomie pathologique de l'acromegalie," *Arch. méd. exper. anat. path.*, **3**, 539–565.

Rasmussen, A. T. ('24). "A Quantitative Study of the Human Hypophysis Cerebri or Pituitary Body," *Endocrinology*, **8**, 509–24.

Smith, P. E. ('27). "The Induction of Precocious Sexual Maturity by Pituitary Homeotransplants," *Amer. J. Physiol.*, **80**, 114–25.

Uhlenhuth, E. ('21). "Experimental Production of Gigantism by Feeding the Anterior Lobe of the Hypophysis," *J. Gen. Physiol.*, **3**, 347–66.

* British 1851 Exhibition Scholar from Canada.

[1] Unless otherwise specified, the terms "giant" and "dwarf" will be used to denote relative differences in adult size, without causal or pathological implications.

[2] These data are not presented in detail.

ON THE CONTOURS OF STELLAR ABSORPTION LINES, AND THE COMPOSITION OF STELLAR ATMOSPHERES

BY CECILIA H. PAYNE

HARVARD COLLEGE OBSERVATORY, CAMBRIDGE, MASS.

Communicated April 13, 1928

The paramount importance of line contour in the analysis of the outer layers of stars has long been recognized, and the contours of strong lines were roughly measured by Schwarzschild[1] and Bottlinger,[2] and studied with more precise methods by Shapley,[3] the writer,[4] Kohlschütter,[5] and von Klüber;[6] they are also coming into prominence in discussions of many kinds, as is shown by the work of Sanford[7] and Curtiss.[8] The theory by which line intensities have been interpreted has not until recently taken account of the shapes of spectral lines, but the work of Unsöld,[9] discussing the formation of absorption lines by a mechanism similar to that considered by Stewart,[10] is in a form suitable for immediate astrophysical application.

This is not the place to discuss the theory developed by Unsöld; Milne[11] has rightly pointed out that it does not apply to the high level chromosphere, but to layers lying closer to the photosphere. It provides a physical picture—doubtless too simple—of the formation of absorption

lines, and the ultimate test of such a theory is its agreement with observation. Accordingly, the writer has selected a few typical spectra from the Harvard collection and analyzed the contours of prominent lines for the purpose of comparison with Unsöld's formula.

The equations derived by Unsöld are obtained by classical methods, the atom being regarded as an isotropic harmonic oscillator. In his earlier paper he deduces for the emergent radiation $b(0,i)$, in terms of the photospheric radiation B, the expression

$$\frac{b(0,i)}{B} = \frac{0.5 + \cos i}{1 + \sigma H} + e^{-\sigma H \sec i} \frac{0.5 - \cos i}{1 + \sigma H}$$

where i is the angle made by the line of sight with the normal to the photosphere, H the height of the atmosphere, supposed homogeneous, and σ the scattering coefficient, which is the essential feature of the theory, and is given by the expression

$$\sigma = \frac{2e^4\lambda_0^2 Nf}{3mc(\lambda - \lambda_0)^2}.$$

Here e = electronic charge
m = mass of electron
c = velocity of light
λ_0 = wave-length of resonance line
λ = wave-length considered
N = number of atoms per cubic centimeter
f = oscillatory strength = $^2/_3$ and $^1/_3$, respectively, for H and K.

The product NH of the number of atoms in a cubic centimeter by the height of the "homogeneous atmosphere" gives the total number of atoms of the kind in question above the photosphere. The theory in its present form applies only to resonance lines.

The formula was designed by Unsöld for center-limb comparisons in the solar spectrum, and in his second paper a more detailed expression, depending on the observed solar law of darkening in different wave-lengths, is derived. In the absence of the corresponding stellar data, the earlier formula is used in the present note, the value $^1/_2$ being used for cos i, so that the second term in the equation for b disappears. With any plausible value for cos i this second term is very small.

The line contours given by the formula always lead to complete blackness at the center, a condition which is never observationally fulfilled. Unsöld attributes the disagreement tentatively to local deviations from monochromatic radiative equilibrium. It has recently been shown by Carroll[12] that, as a result of the finite resolving power of all instruments,

no line can ever *appear* black at the center, even if it actually is so. However, it is evident when we study the contours of lines for various spectral classes that the deviations from central blackness differ greatly from class to class, so that they are not entirely or even primarily instrumental. The deviation has been repeatedly studied, and is discussed by Davidovich[13] and by the writer and Hogg[14] under the name of "incipient central emission."

With spectra of moderate dispersion the contours of only the strongest lines can be studied, and the present note deals with the H and K lines of ionized calcium and the Balmer lines. Even from inspection it is evident that the contours of the hydrogen and calcium lines differ from one another, and that they alter from class to class, and from supergiant to dwarf within the same spectral class. The hydrogen lines are flattest at their maximum of intensity, sharpening in form toward class *M*, and the calcium lines become progressively flatter in shape toward the cool stars. The lines of the supergiants are sharper than those of less luminous stars. Some such flattening with increase in *NH*, the number of active atoms, would be expected from Unsöld's theory, which is graphically illustrated in figure 1.

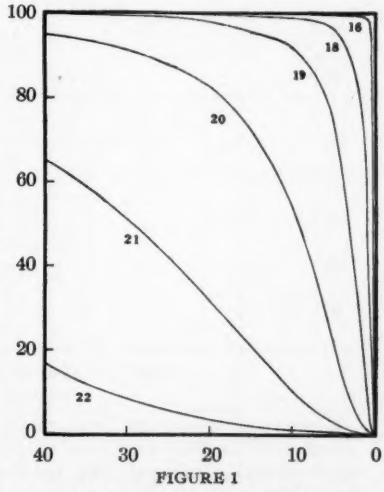

FIGURE 1

Contours of lines given by various numbers of effective atoms in the reversing layer, computed from Unsöld's formula. Ordinates and abscissae are percentage light losses (as compared with the continuous background) and Ångstrom units, measured from the line center. The logarithms of the numbers of effective atoms corresponding to the various contours are shown on the diagram.

It may be seen from the diagram that a number of active atoms less than 10^{15} per square centimeter above the photosphere would not be detected as a spectrum line, and no line (excepting for hydrogen) has been observed with a width corresponding to more than 10^{20} atoms. There should therefore be about 10^5 times as many atoms of Ca^+ at the maximum of the H and K lines as at their marginal appearance. This factor is of the same order as the corresponding ratio derived for the H and K lines from the Fowler-Milne equations, using a partial electron pressure of 10^{-6} atmospheres. As the general agreement of the theory with the phenomena of the spectral sequence is

satisfactory we proceed to examine the individual line contours, and apply the theory to them.

The contours to be discussed were determined from microphotometer tracings of plates made with one, two and three objective prisms. They were reduced by the method described by Hogg,[15] using his standard reduction curve. The consistency of the results from different plates of the same star, shown in figure 2, strongly recommends the method.

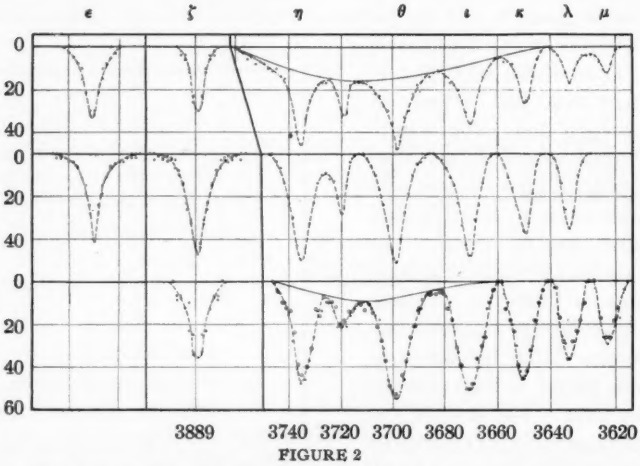

FIGURE 2

Contours of lines for the three stars β Crucis (B1), α Pavonis (B3), and α Eridani (B5). Wave-lengths and Balmer lines are indicated on the lower and upper margins. Percentage light losses are indicated along the left margin. Broken lines represent observed contours in this and the following diagrams. Crosses, throughout the paper, represent reflected points, the lines being assumed to be symmetrical. Circles and dots for the contours of α Eridani represent determinations from two plates; no scale adjustment has been made. The depression of the continuous background, mentioned in the text, is drawn in as a smooth curve.

The contours of the Balmer lines for three *B* stars are shown in figure 2. The theoretical curve does not satisfactorily represent them, even for lines whose wings do not interlace; the lines are too widely spread and their centers are too shallow, a phenomenon even more pronounced for absorption *O* stars than for class *B*. It is notable that the greatest deviations from the theoretical curve are found at the ends of the spectral sequence—at class *B* for the Balmer lines, and at class *M* for the H and K lines.

The lines are measured from the apparent continuous background, which is, of course, depressed to the violet of Hζ by the confluence of the

line wings.[16] The dip in the background that is shown by β Crucis and α Eridani, but not by α Pavonis, is not, however, the result of this wing confluence, which produces a uniform depression; it appears to represent

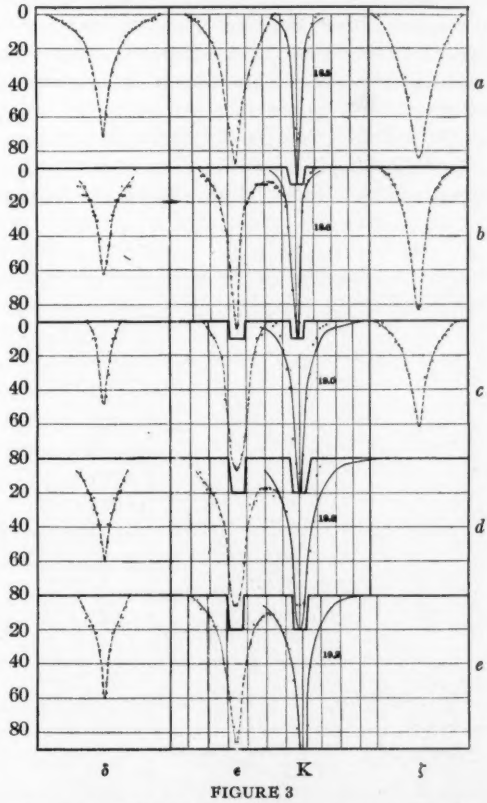

FIGURE 3

Contours of Hδ, Hε, K and Hζ for the stars (*a*) β Pavonis (A5), (*b*) Canopus (F0?), (*c*) θ Scorpii (F0), (*d*) η Scorpii (F2), and (*e*) ρ Puppis (F8). Abscissa and ordinate are wave-length and percentage light loss. The smooth contours are computed from Unsöld's formula for the K line; the logarithm of the number of effective atoms to which they correspond is indicated in the diagram. The broken lines represent the observed contours.

a definite band absorption. It is conspicuous in α Cygni, β Orionis and the Pleiades, and seems to have a maximum near 3820. Apparently it is strong for many stars that show the "cometary cyanogen" depression discussed by Shapley.[17]

In figure 3 the data are tabulated for five stars, ranging in spectral class from $A5$ to $F8$. In addition to the H and K lines, two hydrogen lines are shown, to furnish an estimate of the extent to which Hϵ contributes to the H line of Ca^+. As the contribution is considerable the contour is at present examined only for K; a discussion of the hydrogen lines is reserved for a later note. The H and K lines of ten stars, examined with one-prism dispersion, are shown in figure 4; with one exception the stars are so cool that the hydrogen is no longer the predominating influence at H, and the contours of the two lines may be compared with theory. It should be noted that they are satisfactorily represented in their appropriate widths (ratio 1:1.4). Three special comments may be made.

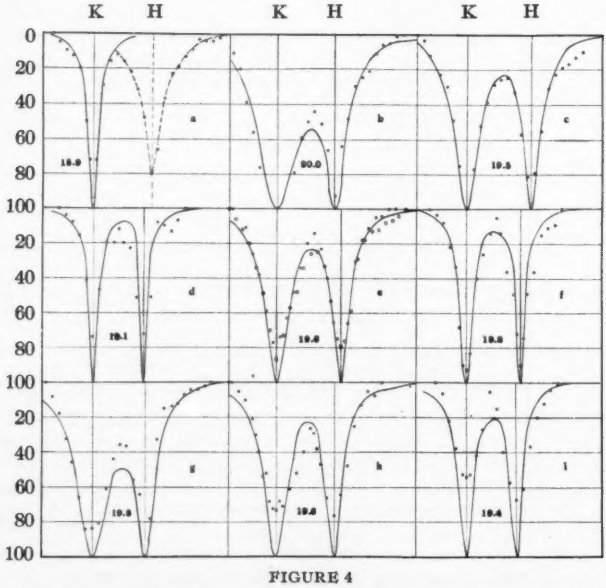

FIGURE 4

Contours of H and K, and theoretical curves, for (*a*) α Circini (A8), (*b*) x Carinae (F8), (*c*) α_1 Capricorni (G5), (*d*) τ Ceti (G9), (*e*) $-52°$ 7028 (dots) and γ Sagittarii (circles), both of Class K0, (*f*) α Arietis (K2), (*g*) $+17°$ 712 (K5), (*h*) ϵ Indi (K5), (*i*) α Orionis (M0).

The star α Circini, whose brightness was discussed by Shapley,[18] appears to belong, on the evidence of its hydrogen contours, to an earlier class than $F0$; it is here reclassified as of class $A8$. On similar grounds Canopus appears to be an even earlier A star with strong metallic lines; it is now placed in class $A7$, and the strength of the lines of ionized calcium is regarded as an index of great brightness.

The widest calcium lines, and the greatest derived value of *NH*, occur for the supergiant *x* Carinae. The large number of atoms above the photosphere is an evident accompaniment of great brightness, arising from the same cause as the systematic strengthening of the lines of bright stars.[19] Lines of the same width and depth as those of *x* Carinae are shown by the Cepheid variable *l* Carinae (median absolute magnitude −3.5); this and other Cepheid variables will be spectroscopically discussed in greater detail in a later paper, and it is mentioned here only to indicate roughly the order of brightness of *x* Carinae. The wider lines are the most sensitive indices of the number of atoms *NH*, and it may be that the contours of H and K will be useful in detecting very bright stars of classes *F* and *G*, whose absolute magnitudes are indeterminate by other methods.

The two coolest stars shown in figure 4 are ϵ Indi and α Orionis—the former a dwarf, the latter a supergiant. Both show a weakening and flattening of the lines, especially of the K line, and as their difference in brightness is so great one is inclined to associate the phenomenon with their low temperature. The data for ϵ Indi are derived from two accordant plates, and the contour of the lines of α Orionis is substantiated by unpublished work of Hogg[20] for this and other stars of class *M*. In discussing the cause of the central deviations of lines from blackness, the shallowness of those of the cool stars, both dwarf and supergiant, is of obvious importance. Evidently the suggestion that central intensity is a linear function of surface gravity for the H and K lines is not substantiated by our data.[21]

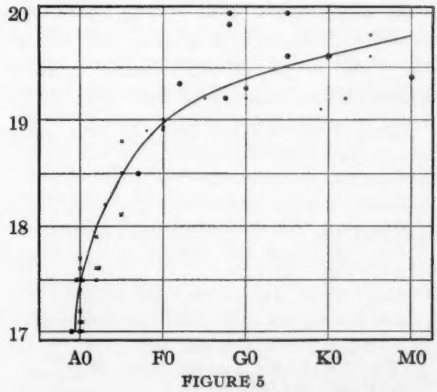

FIGURE 5

Relation between logarithm of number of effective atoms of ionized calcium (ordinate) and spectral class (abscissa). Large, medium and small dots represent supergiants, giants and dwarfs. The value for the sun, derived by Unsöld, is marked with a circle. Values derived from unpublished data by Dunham are indicated by crosses.

In conclusion we summarize in figure 5 the data derived in this paper for the number of calcium atoms in the ionized state above the photosphere. There is a progression in *NH* with spectral class, and for class *M* the number of active atoms begins again to decline. The diagram gives

us a picture of the progression in ionization and also of the different amounts of atmosphere present for stars of about the same temperature and differing luminosity. It should be noted that the flat maximum of the Fowler-Milne formula[22] does not appear.

The method outlined by Unsöld permits us virtually to weigh the atmospheres of individual stars. The process has here been performed for calcium, but with greater resolving power it would be possible to carry out for many stars the detailed analysis made by Unsöld for the sun. As a method of obtaining comparative data about the atmospheres of individual stars, his theory has considerable value. One of its great advantages is that it deals with the wings of lines, and is therefore independent of saturation.

In a recent paper Fairley[23] makes an estimate, with the aid of Stewart's formula,[10] of the number of sodium atoms in the solar atmosphere. His value of 4×10^{14} differs from that of Unsöld (2.6×10^{16}) for several reasons, the chief being the difference in what the two investigators regard as the width of the line. Using Unsöld's more accurate line contours for the solar D lines, and a value of 0.99 for Fairley's "contrast factor," $1/(1 + k)$, we should obtain from Stewart's formula a number of sodium atoms in close accord with that derived by Unsöld.

[1] *Sitz. d. Preuss. Akad. d. Wiss.*, **46**, 1914 (1183).
[2] *Ast. Nach.*, **195**, 1913 (117).
[3] *Harv. Obs. Bul.*, 805, 1924.
[4] *Harv. Obs. Repr.*, 28, 1926.
[5] *Ast. Nach.*, **220**, 1924 (326).
[6] *Zeit. f. Phys.*, **44**, 1927 (481).
[7] *Ap. J.*, **66**, 1927 (170).
[8] *Mon. Not. R. Ast. Soc.*, **88**, 1928 (205).
[9] *Zeit. f. Phys.*, **44**, 1927 (793); **46**, 1928 (765).
[10] *Astrophys. Journ.*, **59**, 1924 (30).
[11] *Mon. Not. R. Ast. Soc.*, **88**, 1928 (188).
[12] *Ibid.*, **88**, 1927 (154).
[13] *Harv. Obs. Bul.*, 846, 1927.
[14] *Harv. Obs. Circ.*, 303, 1927.
[15] *Harv. Obs. Bul.*, 855, 1928.
[16] *Harv. Obs. Repr.*, 46, 1928.
[17] *Harv. Obs. Bul.*, 856, 857, 1928.
[18] *Ibid.*, 798, 1924.
[19] *Harv. Obs. Circ.*, 307, 1927.
[20] Cf. *Pop. Ast.*, **36**, 1928 (236).
[21] Cf. Eddington, *The Internal Constitution of the Stars*, 1926, p. 341.
[22] *Mon. Not. R. Ast. Soc.*, **84**, 1924 (299).
[23] *Astrophys. Journ.*, **67**, 1928 (114).

REGIONAL ISOSTATIC REDUCTION OF GRAVITY DETERMINATIONS

BY GEORGE R. PUTNAM

U. S. DEPARTMENT OF COMMERCE, WASHINGTON, D. C.

Communicated April 11, 1928

The isostatic compensation of the topographic features of the earth's crust must be regional to some extent. Considering the physical characteristics of the materials of the crust, perfect local compensation is inconceivable, as it would require that each unit prism of small cross-section should be capable of moving vertically, free from friction with any adjoining prism. The known strength of the rock materials of the crust makes it certain that there are appreciable departures locally from such a condition, even though the general approach to perfect equilibrium is close, as it is believed to be.

It is important to ascertain whether the departures from complete local isostasy are sufficiently large to warrant their being taken into account in the reduction of gravity determinations. The best method of reduction will be that one which yields anomalies most nearly fitting the data as derived from gravity determinations or other evidence. Although the facts sought are difficult to sift out because of the great number of possible variants in the conditions of the crust, there is good evidence that the departures from complete local compensation are sufficient in their effect as shown in anomalies of gravity to require that they be considered in a complete and systematic reduction of gravity observations. This evidence is summarized below, under three heads, *a*, *b* and *c*. It is based on the fact that in a regional, as compared with a local distribution of compensation, the amount of compensation beneath the zones in the vicinity of the station will be less for stations above the general level, and will be more for stations below the general level; and the anomalies for the "above" stations will be smaller algebraically, and for the "below" stations will be larger, than the local compensation anomalies; there may be an opposite effect in remoter zones but this, on account of greater distance, will only partially offset the effect in the near zones. The Hayford method is here taken as illustrative of complete local compensation, as it uses zones of outer radius of 2 meters, 68 meters, 230 meters, etc., with 11 zones within a radius of 12 miles from the station, and computes the compensation effect on the basis of perfect local compensation for each compartment of each zone.

(*a*) Comparison of the anomalies for pairs of stations close to each other horizontally, but of considerable difference of elevation, has the advantage of largely eliminating distant effects and other sources of error.

Results are available for 16 such pairs,[1] with height differences from 531 to 2452 meters. With one minor exception, these show in every case a positive difference when the Hayford anomaly for the lower station is subtracted from that for the higher,[2] the average difference being + 0.015 dyne. This difference is of the sign to be expected from an error due to using complete local compensation, and is evidence of its existence. It is of a size appreciable in gravity reductions, being about the size of the average gravity anomaly. If the difference is due to this cause, and assuming, as is indicated, that it is related to the difference in elevation, it corresponds to an average error in high mountainous regions of +0.0014 dyne for each 100 meters greater elevation. There is one additional pair, Mauna Kea-Honolulu, not included in this group, because of the considerable horizontal distance between the stations, and the less complete data. This pair of stations is of very special importance in this problem, however, because the difference in elevation is far in excess of that of any other pair, being 3975 meters (13,041 feet), and because it brings in oceanic island stations. The anomaly difference for this pair is of the same sign as the others, and it is further evidence of error due to the application of complete local compensation.

(*b*) For stations in mountainous regions, the Hayford anomalies have been grouped for stations above and for stations below the general elevation within 100 miles (Table A), in the order of the differences from the general level, and for these stations there have been computed also the anomalies with compensation based on the assumption that the surface about each station is leveled off to a radius of 37 miles from the station, this area being designated zone *M*, and this has also been done for a radius of 104 miles, designated as zone O^3 (columns 6 and 7, Table A). These latter represent an approximate application of regional compensation, but for distinction they are here designated as "average level anomalies." Comparing the zone *M* or the zone *O* anomalies (columns 6 and 7) with the Hayford local anomalies (column 4) it is found that in nearly all cases the local anomaly is algebraically larger when the station is above the general level, approximately in proportion to the elevation difference, and is smaller in similar proportion when the station is below the general level; in 44 comparisons with zone *M* anomalies there are only 6 which as to sign are at variance with this (column 8). As the local anomaly should be the larger for "above" stations, and the smaller for "below" stations, it is reasonable to ascribe these systematic differences also to error due to a reduction based on complete local compensation.

A correction to the Hayford local anomalies for error due to local compensation is shown to be necessary by the comparisons of pairs as described under (*a*), and by the systematic differences just mentioned, and is also indicated by the large size of several of the high mountain local anomalies.

The factor of 0.0014 dyne for each 100 meters above the general level, found under (*a*), is derived from pairs of stations in very mountainous regions, often with one of the stations near a high summit. Consideration of the distribution of the load of high mountains makes it reasonable to assume that this factor varies in some relation to the height above the general level, or in other words that the difference between the local anomalies and the correct regional anomalies increases by some function of the elevation difference greater than a simple proportion. Inspection of the anomalies indicates this to be true. For the present examination of this subject it is considered sufficient to use the factor 0.0014 dyne per 100 meters for stations from 1000 to 2500 meters above the general level, and to halve this factor for stations of smaller difference of elevation, and to double it for the one station which is of much greater difference of elevation.

The corrections to the Hayford anomalies on account of local compensation, and the factors used in obtaining these corrections as above explained, are shown in table A, column 3. These corrections are positive for stations below the general level, and negative for stations above. The corrected anomalies thus obtained are designated "regional anomalies" (column 5). Comparing these regional anomalies with the Hayford local anomalies, the mean with regard to sign is reduced from +0.008 dyne to +0.002, and the mean regardless of sign is reduced from 0.022 to 0.018. Subtracting either the zone *M* or the zone *O* anomalies from the regional anomalies, the results are more accordant than are the corresponding differences from the local anomalies. In columns 9 and 8 are given these differences, respectively, for zone *M*, excepting that zone *O* is used for the 6 "high" stations for reasons that will be explained later. For "above" stations, the mean difference with regard to sign is reduced from +0.015 local to 0.000 dyne regional, and for "below" stations from −0.003 local to 0.000 regional. The marked systematic tendency of the signs of these differences with the Hayford local anomalies disappears with the regional anomalies.

The sign of the corrections applied here to the Hayford local anomalies corresponds to that which would be caused by errors due to complete local compensation, and the results become more consistent in a number of ways when this is corrected for. The four high mountain stations, including Honolulu which is on a mountain mostly submerged, show marked improvement in anomalies with regional compensation. All of this is therefore substantial evidence against complete local compensation.

(*c*) For five of the pairs of adjacent stations previously referred to there are available[4] comparative anomalies from computations on the basis of complete local compensation, and of average level compensation to three different zones, and table B shows the pair differences of these comparative anomalies, arranged in order of difference of elevation; also in column

TABLE A

COMPARISON OF REGIONAL, WITH LOCAL AND AVERAGE LEVEL ANOMALIES, FOR STATIONS IN MOUNTAINOUS REGIONS

(1) STATION	(2) AVERAGE ELEVATION WITHIN 100 MILES, MINUS STATION ELEVATION. 100 METERS	(3) CORRECTION TO LOCAL TO OBTAIN REGIONAL ANOMALY 0.0007 × (2) DYNES	(4) LOCAL ANOMALY (HAYFORD) DYNES	(5) REGIONAL ANOMALY (4) + (3) DYNES	(6) AVERAGE LEVEL ANOMALY ZONE *M* (37 MILES) DYNES	(7) AVERAGE LEVEL ANOMALY ZONE *O* (104 MILES) DYNES	(8) LOCAL MINUS ZONE *M* ANOMALY (4) − (6) DYNES	(9) REGIONAL MINUS ZONE *M* ANOMALY (5) − (6) DYNES
Stations *below* general level								
Gallup, New Mex.	+0.3	0.000	−0.013	−0.013	−0.013	−0.020	0.000	0.000
State College, Penn.	+0.3	0.000	−0.021	−0.021	−0.019	−0.021	−0.002	−0.002
Goldfield, Nev.	+1.1	0.000	−0.013	−0.013	−0.012	−0.009	−0.001	−0.001
North Hero, Vt.	+1.7	+0.001	+0.001	+0.002	+0.005	+0.005	−0.004	−0.003
El Paso, Tex.	+2.0	+0.001	+0.007	+0.008	+0.008	+0.013	−0.001	0.000
Heppner, Ore.	+2.6	+0.002	−0.027	−0.025	−0.025	−0.016	−0.002	0.000
Olympia, Wash.	+3.1	+0.002	+0.033	+0.035	+0.034	+0.044	−0.001	+0.001
Boulder, Mont.	+3.1	+0.002	−0.015	−0.013	−0.018	−0.013	+0.003	+0.005
Skyhomish, Wash.	+3.2	+0.002	−0.028	−0.026	−0.028	−0.039	0.000	+0.002
Guernsey, Wyo.	+3.2	+0.002	+0.036	+0.038	+0.041	+0.046	−0.005	−0.003
Winnemucca, Nev.	+3.5	+0.002	−0.009	−0.007	−0.006	+0.003	−0.003	−0.001
Sheridan, Wyo.	+3.8	+0.003	+0.032	+0.035	+0.041	+0.030	−0.000	−0.006
Rock Springs, Wyo.	+3.8	+0.003	+0.013	+0.016	+0.013	+0.021	0.000	+0.003
Gunnison, Col.	+3.8	+0.003	+0.020	+0.023	+0.028	+0.018	−0.008	−0.005
Colorado Springs, Col.	+4.2	+0.003	−0.007	−0.004	−0.008	−0.003	+0.001	+0.004
Salt Lake City, Utah	+5.7	+0.004	+0.010	+0.014	+0.013	+0.016	−0.003	+0.001
Denver, Col.	+5.7	+0.004	−0.016	−0.012	−0.007	+0.001	−0.009	−0.005
Boise, Idaho	+5.8	+0.004	+0.003	+0.012	+0.012	+0.022	−0.004	0.000
Sandpoint, Idaho	+5.9	+0.004	+0.002	+0.006	+0.006	+0.011	+0.004	0.000
St. Maurice, Switzerland	+6.5	+0.005	+0.003	+0.008	+0.008	−0.014	+0.005	0.000
Grand Canyon, Ariz.	+8.2	+0.006	−0.010	−0.004	−0.009	−0.019	−0.001	+0.005
Grand Junction, Col.	+8.5	+0.006	+0.024	+0.030	+0.031	+0.038	−0.007	−0.001
Green River, Utah	+8.7	+0.006	−0.021	−0.015	−0.014	−0.001	−0.007	−0.001
For 23 "below" stations								
Mean with regard to sign			0.000	+0.003	+0.004	+0.005	−0.003	0.002
Mean regardless of sign			0.016	0.017	0.017	0.019	0.003	0.000

Stations *above* general level		0.0007 × (2)						
Las Vegas, New Mex.	−0.2	0.000	+0.003	+0.003	+0.003	−0.007	0.000	0.000
Ely, Nev.	−0.2	0.000	−0.021	−0.021	−0.022	−0.030	+0.001	+0.001
Helenwood, Tenn.	−0.3	0.000	+0.040	+0.040	+0.040	+0.037	0.000	0.000
Lower Geyser Basin, Wyo.	−0.6	0.000	−0.001	−0.001	0.000	−0.009	−0.001	−0.001
Norris Geyser Basin, Wyo.	−1.4	−0.001	+0.021	+0.020	+0.022	+0.008	−0.001	−0.002
Pleasant Valley Jct., Utah	−1.5	−0.001	+0.004	+0.003	+0.001	−0.008	+0.003	+0.002
Grand Canyon, Wyo.	−2.5	−0.002	−0.002	−0.004	−0.001	−0.017	0.001	−0.003
Alpine, Texas	−2.6	−0.002	+0.021	+0.019	+0.023	+0.008	−0.002	−0.004
Nogales, Ariz.	−2.9	−0.002	−0.050	−0.052	−0.055	−0.057	+0.005	+0.003
Deer Park, Md.	−2.9	−0.002	+0.010	+0.008	+0.007	+0.001	+0.003	+0.001
Lake Placid, N. Y.	−3.1	−0.002	+0.006	+0.004	+0.003	−0.007	+0.003	+0.001
Hughes, Tenn.	−4.3	−0.003	−0.029	−0.032	−0.033	−0.042	+0.004	+0.001
Lead, S. Dak.	−4.7	−0.003	+0.052	+0.049	+0.049	+0.039	+0.003	0.000
Yavapai, Ariz.	−5.1	−0.004	+0.001	−0.003	+0.001	−0.007	0.000	−0.004
Truckee, Cal.	−5.1	−0.004	−0.028	−0.032	−0.032	−0.057	+0.004	0.000
Stations of great elevation above general level		0.0014 × (2)						
							Zone *O*	Zone *O*
Mt. Hamilton, Cal.	−12.0	−0.017	−0.003	−0.020	−0.011	−0.018	+0.015	−0.002
Cloudland, Tenn.	−13.2	−0.018	+0.004	−0.014	−0.002	−0.011	+0.015	−0.003
Honolulu, H. I.	−19.4	−0.027	+0.052	+0.025	+0.044	−0.013	+0.065	+0.038
Gornergrat, Switzerland	−20.2	−0.028	+0.049	+0.021	+0.031	+0.002	+0.047	+0.019
Pike's Peak, Col.	−20.3	−0.028	+0.021	−0.007	+0.008	+0.004	+0.017	−0.011
		0.0028 × (2)						
Mauna Kea, H. I.	−56.5	−0.158	+0.183	+0.025	+0.149	+0.055	+0.128	−0.030
For 21 "above" stations								
Mean with regard to sign			+0.016	+0.001	+0.011	−0.006	+0.015	0.000
Mean regardless of sign			0.029	0.019	0.026	0.021	0.015	0.006
For all 44 stations								
Mean with regard to sign			+0.008	+0.002	+0.007	0.000	+0.005	0.000
Mean regardless of sign			0.022	0.018	0.021	0.020	0.009	0.004

5 the regional anomaly differences are shown as derived by applying corrections to the Hayford local anomalies in the manner just described (and given in Table A, columns 3 and 5).

The results for the mountain pairs systematically favor regional compensation, or average level compensation, as compared with local compensation. This is particularly true for the 3 pairs with great differences of elevation. The last pair, in Arizona, in which the upper station is on a plateau and the lower station in a relatively narrow gorge, shows practically no difference in the local and average level anomalies; this result is consistent with the others, as for such locations there is no large local load to be distributed, and there should be little error due to applying local compensation.

These results show much greater consistency with regional than with local compensation, the reduction in the mean of the differences being from +0.050 to +0.006; also each of the differences is much smaller, and they are no longer all positive in sign. Although few in number, the results from these pairs are strong evidence for regional as against local compensation, because they cover extreme conditions and are consistent, and because of the strength of this method of comparing pairs of nearby stations.

TABLE B

COMPARISON OF GRAVITY ANOMALIES FROM PAIRS OF STATIONS, WITH RESPECT TO LOCAL OR REGIONAL COMPENSATION

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			DIFFERENCE IN ANOMALY, SUBTRACTING ANOMALY FOR LOWER FROM HIGHER STATION				
					AVERAGE LEVEL COMPENSATION		
STATION	ELEVATION METERS	DIFFERENCE IN ELEVATION METERS	LOCAL COMPENSATION DYNE	REGIONAL COMPENSATION DYNE	TO 12 MILE ZONE *K* DYNE	TO 37 MILES ZONE *M* DYNE	TO 104 MILES ZONE *O* DYNE
Pairs with higher station on mountain summit:							
Mauna Kea, H. I.	3891	3975	+0.131	0.000	+0.118	+0.105	+0.068
Honolulu, H. I.	6						
Gornergrat, Switzerland	3015	2597	+0.046	+0.013	+0.041	+0.032	+0.016
St. Maurice, Switzerland	419						
Pike's Peak, Col.	4293	2452	+0.028	−0.003	+0.020	+0.016	+0.012
Colorado Springs, Col.	1841						
Cloudland, Tenn.	1890	896	+0.033	+0.018	+0.031	+0.031	+0.031
Hughes, Tenn.	994						
Pair with higher station on plateau:							
Yavapai, Ariz.	2179	1330	+0.011	+0.001	+0.010	+0.010	+0.012
Grand Canyon, Ariz.	849						
For 5 pairs:							
Mean with regard to sign			+0.050	+0.006	+0.044	+0.037	+0.028
Mean regardless of sign			0.050	0.007	0.044	0.037	0.028

The foregoing evidence shows that a regional reduction of gravity

determinations gives results nearer the truth than a complete local compensation. In uneven regions the compensation cannot be perfect under each small area separately, nor can it be uniform under any considerable area. It must lie somewhere between these, and can be represented as the compensation required by topography limited by a warped surface which for summit areas would be below the earth's surface but above the average plain, and for valley areas would be above the surface but below the average plain (Fig. 1). This lack of perfect local compensation simply represents the resistance of crustal materials to adjustment; it is a sort of lag, in which the time element is undoubtedly also quite important.

There are other explanations offered of the systematic discordances noted in the gravity anomalies with the local compensation method, but these are inadequate or improbable, and the fact here pointed out, that there is actually a moderate degree of regional effect in compensation which cannot be ignored, is a simple and logical explanation of the observed phenomena, and fits the data consistently. The writer in 1909 suggested this source of error with local compensation.

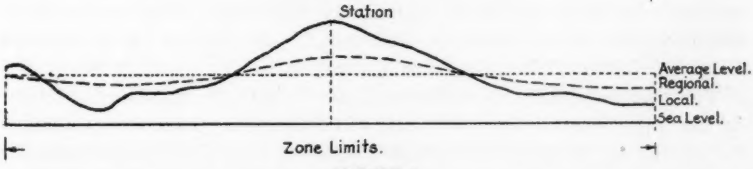

FIGURE 1

Diagram illustrating REGIONAL (– – – – – –), LOCAL (————), and AVERAGE LEVEL (------) Compensations. The diagram illustrates in cross-section, exaggerated and not drawn to scale, the actual topography above sea level, corresponding to local compensation, and the rearrangement of this amount of topography to illustrate regional compensation, or average level compensation, within a limited zone (as to 100 miles). The actual compensation, which is a deficiency of density extending a considerable distance below, is not represented.

To illustrate the idea, consider an outstanding mountain such as Pike's Peak. The Hayford local compensation method takes the core of that mountain, a cylindrical column 2 meters in radius and 70 miles deep, and computes the compensation as though the deficiency of density would exactly balance the portion above sea level in that column independently. If the top of the column were removed, to continue this state of perfect equilibrium, it would be necessary for this column in some manner to readjust itself regardless of friction from the surrounding materials. The same would be true of the next zone, which is a cylindrical shell, with 2 meters inner radius and 68 meters outer radius, and also 70 miles long; and so on for other zones. The mere statement of this hypothesis indicates the probability that it departs sufficiently from

actual possibilities in nature to affect the correctness in mountainous regions of gravity reductions using complete local compensation; its use is not called for by mathematical convenience, as a partial regional compensation involves less labor.

The following is a plan for the practical application of limited regional compensation in the reduction of gravity observations. This plan is based on the fact shown in both tables A and B, that for the high mountain stations the regional anomalies accord better with the zone *O* than with the zone *M* anomalies, and that for the remaining 38 stations the zone *M* anomalies are quite close to the regional anomalies, the average difference being only 0.001. Until some practicable and more accurate method of computing the actual regional compensation is developed, a close approximation will be obtained by using a single zone computation of compensation to the limit of zone *O* for mountain stations more than 1000 meters above the surrounding country, and the same to the limit of zone *M* for the other stations, in addition to the Hayford zone computation beyond in each case. The results will be less subject to systematic error in mountainous regions than the results with complete local compensation. For stations not in mountainous regions the differences between the local and the zone *M* anomalies are quite small; the average for 82 stations is 0.001, and no difference exceeds 0.004 dyne.

The reason for using two different zones is this. As will be shown later in this paper, the evidence indicates that there is regional compensation, to some extent at least, extending to a distance of 100 miles from a station; such regional compensation will be most apparent in the anomalies for high mountain stations, and less apparent for stations in other regions. For the latter there appear to be other uncertainties arising from using an average level compensation for such a large radius as 104 miles, which more than overcome the advantage of taking the compensation as so widely regional.

As the computation of the compensation for a single zone, to the limit of zone *O*, 166 kilometers (104 miles), instead of for 15 zones as in the Hayford method, is a material saving in labor, it would be a further saving to compute as for a single zone the attraction for a considerable area, and this can be done with accuracy to the limit of zone *L*, within which the effect of curvature is inappreciable. The direct effect of departures from a plain surface of the topography surrounding a station can be computed better separately. It is also possible to eliminate the necessity of computation for individual stations, of the attraction and compensation for the area of the earth beyond a distance of 1429 miles from the station (zones 7 to 1);[5] these values could conveniently be tabulated for various grand divisions of the earth. The foregoing should make it possible without loss of accuracy to reduce materially the number of zones.

The Hayford method in its systematic treatment of gravity determinations much surpassed any preceding work, but some misconception as to isostasy would have been avoided and labor saved if earlier procedure had not been so wholly discarded. Some regional treatment was rather clearly indicated. Incidentally, this would reduce the great number, 317 for each station, of separately computed corrections to gravity, and the consequent likelihood of small cumulative errors. The Hayford method was especially valuable in computing, for the first time, the effect of the intermediate zones to 1500 miles distance; for these zones of large area it is really a regional reduction.

To summarize, the following "regional isostatic reduction" of gravity observations is proposed: this designation is used in order to distinguish from the several other isostatic methods. (1) Compute the attraction as that of a horizontal plate of average density, of thickness equal to the elevation of the station above sea level, and with 28,800 meters radius (limit of zone *L*). (2) Compute the effect of the departure of the surface about the station from a plain as accurately as need be, by the well-known formulas. (3) Compute the compensation in one operation for the average elevation, within a radius of 166,700 meters (zone *O*) for mountain stations over 1000 meters above the average level, and within a radius of 58,800 meters (zone *M*) for other stations. The average elevation of the surface within these limits, respectively, will be obtained by dividing the area into equal compartments. (4) Take the attraction and compensation corrections from the Coast and Geodetic Survey tables for the zones beyond the limits stated under (1) for the attraction, and under (3) for the compensation, and extending through zone 8. (5) For the area beyond zone 8, use values tabulated for grand divisions. (6) Combine (1) to (5) to obtain the total effects of attraction and compensation, stated separately. (7) Compute the correction for elevation by the known formula. (8) Apply (6) and (7) to the observed value of gravity to obtain the observed value reduced to sea level; the reduction to sea level, instead of the reverse, is preferable for several reasons.

A substantial application of this method to the 44 stations in table A, using available computation data, gives these results for regional anomalies; mean with regard to sign +0.002 dyne, mean regardless of sign 0.018; these are the same as the means of the regional anomalies derived by correcting the Hayford local anomalies, and are appreciably smaller than the latter.

Horizontal Extent of Regional Distribution of Compensation.—In a strict sense, there cannot be a definite horizontal limit of regional compensation, as the compensation must vary gradually from one area to another without abrupt changes. The load of a large mountain will be greatest, per unit area, beneath its summit, but this load may be widely distributed, and probably is still appreciable at a considerable distance, as 100 miles.

There will not be any sharp limit at which the load of the mountain, or its compensation, will cease, and there will be no separation between the compensation for this mountain and for every other feature within effective range. It should be noted that a mountain is a real load above the average level, regardless of any reasonable theory as to how the mountain came to be there or how it is compensated.

Tests have been made to determine the probable extent of regional compensation,[6] by comparing the Hayford local anomalies with anomalies computed using uniform compensation sufficient for the average elevation within zones of various radii, up to 104 miles (zone *O*).

This test in the manner made compares two arrangements of compensation, neither of which, strictly speaking, is physically possible. It has been shown in this paper that the hypothesis of complete local compensation leads to appreciable error. But uniform compensation to a distance of 104 miles, while apparently a fairly close approximation, cannot in uneven regions accurately correspond to any actual system of regional isostasy. In mountainous regions regional compensation will be more correctly represented by a warped surface than by a plain. For example, Pike's Peak and Gunnison are 100 miles apart; the average elevation within zone *O* for Pike's Peak is nearly 500 meters higher than the average elevation within zone *O* for Gunnison. These zones overlap by 100 miles, and for this overlapping area, in this test as made, quite different compensations were used for Pike's Peak and for Gunnison. The average differences developed in this test of local versus regional compensation are extremely small, being only 0.001 and 0.002 dyne, the minimum limit used for anomalies. It is probable that the effect sought is masked by effects due to the fact that the regional compensation used is not exact. It is doubtful whether the results thus far obtained by this method furnish valid evidence as between different limits of regional compensation.

A method is available, however, which largely obviates this difficulty. This is the consideration of pairs of gravity stations near each other horizontally, but differing in elevation.[7] An error due to using average elevations about the two stations of such a pair, as well as other possible errors, should affect them approximately to the same degree, and therefore should be eliminated in comparing the anomaly differences. Some results by this method are shown in table B. These show that the regional anomaly differences agree best with, and are fairly accordant with, the average level differences for zone *O*. The only large spread is in the Mauna Kea-Honolulu pair, which involves a much greater mountain mass and greater difference of elevation than any of the other pairs. These results indicate appreciable regional compensation to a radius near 100 miles, and probably larger for so great an individual mountain as Mauna Kea.

The results given in table A disclose another method of investigating the horizontal extent of the regional compensation about a station. If the regional anomalies (column 5) are obtained by an independent method as was done in this case, then these may be compared with average level anomalies computed for zones of different radii. In table A this may be done for zones M (37 miles) and O (104 miles). Considering first the 6 high mountain stations, the average anomalies with regard to sign are, regional +0.005 dyne, zone O +0.003, and zone M +0.036; this and a comparison of individual differences show advantage for zone O.

The evidence from both these methods, though meager, indicates that there is regional compensation effect to some distance approaching 104 miles from the station, or possibly exceeding that distance for great mountains. On the other hand, for the 38 stations other than the high mountains, the mean anomalies with regard to sign are, regional +0.002, zone M +0.002, and zone O −0.002, and in only 3 out of 38 cases is the regional anomaly nearer to the zone O than to the zone M anomaly, showing a marked advantage for zone M. For reasons already explained, these results are not contradictory. It is probable that the regional effect for a high mountain is really appreciable at 100 miles distance, or possibly more. But the effect of such wide regional compensation will be much less apparent for stations less elevated above the average level, and will be somewhat masked by uncertainties due to carrying the average level compensation to such distances.

Summary.—The isostatic compensation of the topographic features of the earth's crust must be regional to some extent. Perfect local compensation is inconceivable; it could only be obtained if the materials were so plastic that no surface irregularities would remain.

The departures from complete local compensation are sufficient in their effect to require that they be considered in a complete reduction of gravity observations. The evidence is stated, and it is consistent.

A reduction using limited regional compensation will give anomalies nearer the truth in mountainous regions than one with a complete local compensation. Such a distribution may be represented as the compensation required by topography bounded by a warped surface lying between the earth's surface and the average level in any region.

A "regional isostatic reduction" method is proposed and outlined; in rugged regions this yields results nearer the truth than the completely local Hayford method, and for similar precision it requires less labor. It is based on a different conception of isostatic compensation, and it accounts for discordances not before explained.

There are uncertainties in the method which has been used for testing the probable extent of regional compensation, which are probably sufficient to make doubtful the validity of results on this subject previously obtained.

An effective method for this purpose is the comparison of pairs of adjacent stations of considerable difference in elevation, a differential method which eliminates various possible sources of error. Some results with this method are given, and they favor the correctness of regional compensation. Another method of investigating regional compensation is illustrated, based on a comparison of regional anomalies obtained by independent methods. There are indications from these two methods of the correctness of an assumption equivalent to regional compensation extending 100 miles from a station, or further for great mountains.

The validity of limited regional compensation affects the discussion of the so-called Airy and Pratt theories. With regional compensation the density will not vary just according to the surface elevation, and bulges or crust thickenings replace the idea of roots.

[1] Bowie, *Coast and Geodetic Survey, Special Publication*, No. **40**, p. 93 (1917).

[2] This fact and the method of comparing stations by pairs were pointed out first by the writer, *Science*, **36**, p. 869 (1912). They are of prime importance and furnish the key to the question of local versus regional compensation.

[3] Hayford and Bowie, *Coast and Geodetic Survey, Special Publication*, No. **10**, p. 100 (1912); No. **40**, pp. 66, 90, 91 (1917). The lettered and numbered zones mentioned here, and in the following text, are those of the Hayford reduction.

[4] Putnam, *Bulletin, Geological Society America*, **33**, p. 301 (1922).

[5] *Coast and Geodetic Survey, Special Publication*, No. **10**, p. 71 (1912).

[6] *Coast and Geodetic Survey, Special Publication*, No. **10**, p. 98 (1912); No. **40**, p. 85 (1917).

The writer suggested this method in 1909; while it does not solve the question of regional compensation, it gives results of interest otherwise, as shown in this paper.

[7] *Bulletin, Geological Society America*, **33**, 299 (1922); *Science*, **36**, 869 (1912).

PROOF OF ISOSTASY BY A SIMPLE GRAVITY REDUCTION METHOD

By George R. Putnam

U. S. Department of Commerce, Washington, D. C.

Communicated April 11, 1928

The equilibrium theory of the crust is so important in earth problems, that it appears worth while to consider an independent proof of isostasy. This paper will give such a proof, and also an account of a short isostatic method for the approximate reduction of gravity observations. This method affords a simple means for the purpose, gives results which are in fair accord with the observed phenomena, and brings out certain general conceptions of isostatic compensation which are somewhat obscured in a more detailed small area treatment.

The systematic gravity reduction method of Hayford published in 1909, and discussed in the preceding paper, took into account for the first time the effect of the curvature of the earth,[1] and this is necessary to a complete analysis.

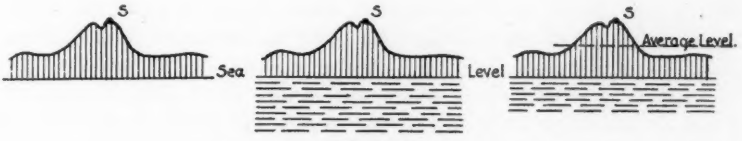

FIGURE 2

For station *above* the average level

1. Bouguer reduction; no compensation; isostasy erroneously ignored.

2. Free-air reduction; plate compensation for station elevation, showing erroneous *over* compensation.

3. Average elevation isostatic reduction (1895); plate compensation for average elevation, approximately correct; compensation is represented as the negative attraction of plate of the same density as the topography.

For station *below* the average level.

1. Bouguer reduction.

2. Free-air reduction; showing erroneous *under* compensation for station below average level.

3. Average elevation reduction.

Diagram illustrating comparison of earlier gravity reduction methods, and the large errors in the first two due to neglect of isostasy, or incorrect compensation, and the correction of these large errors by the average elevation reduction in 1895.

In these examples the average elevation is the same for the two stations. The vertical scale is exaggerated to emphasize the compensation effects. If drawn the same as the horizontal scale, a better idea would be given of the attraction correction. The compensation (shown by horizontal lines) is, for convenience, placed immediately below sea-level, and of thickness equal to the station elevation and the average elevation, respectively.

The larger discrepancies with the early gravity reduction methods were not, however, due to neglect of curvature. These discrepancies were very large, the maximum being as great as +0.628 dyne for the

free-air reduction, and −0.263 for the Bouguer reduction. They were in fact due to neglect of isostasy in the latter, and to erroneous compensation in the former. (Fig. 2.) The extreme anomalies were many times greater than would be accounted for by neglect of curvature. The neglect of all topography and compensation beyond 1430 miles from a station, it has been shown, would introduce a relative error of only about 0.004 dyne, equivalent to the average error of an observation. A zone of this radius includes only one per cent of the surface of the earth, and therefore the neglect of 99 per cent of the earth's surface has a barely appreciable effect in relative gravity reductions. All of the gravity results under consideration are relative only, and it is important to note that relatively they are not affected by any source of error to the extent that the error affects all stations equally.

From the actual results it is difficult to ascertain the magnitude of errors due to neglect of curvature, because these are involved with errors from other sources. In a comparison of the Hayford and free-air anomalies for 25 continental stations in fairly flat regions, the maximum difference is 0.013 dyne, and the average is 0.006. The free-air anomalies are derived with a complete neglect of curvature, while the Hayford method includes curvature. Inspection shows that these small differences have some systematic relations to the average elevation of the immediate vicinity, so that the actual effect of neglect of curvature (and remote topography) is perhaps materially less than the above figures. Theory supports this numerical evidence. Both the attraction and the compensation may be looked upon as spherical shells of the outer portions of the earth, whose variations from spherical symmetry are very small, less than one-one-thousandth part, in comparison with their principal dimensions.

The errors due to neglect of curvature are relatively so small that they do not affect the proof of isostasy from gravity determinations.

It is therefore possible, by developing the simple early method of computation based on the attraction of indefinitely extended horizontal plates, to obtain for any station a fair approximation to the value of gravity reduced to sea level, and from comparisons of such results to obtain proof that the earth's crust is in a condition of equilibrium. This is the method used by the writer in 1895.

The attraction of the topography above sea level, considering the latter as a plain of indefinite extent, may readily be computed by the formula for the attraction of an extended horizontal plate, on a point without or in its surface, the thickness of the plate being the elevation of the station, and a simple correction being added, where appreciable, for the inequalities of the upper surface. For any assumed density, the attraction thus computed will be correct excepting for neglect of the effect of curvature.

Considering the earth's crust to be in a condition of equilibrium, the effect of the compensation necessary to satisfy this condition may also readily be computed as the negative attraction on a point without it, of an extended horizontal plate. The compensation is really a small deficiency of density extending downward for a considerable distance, but in this computation it is convenient and allowable to take it as the equivalent negative attraction of a plate of the same density as the topography.

The important point with this reduction is to determine a thickness for this compensation plate that is reasonable and will satisfy the conditions. The total compensation effect on gravity at any station is the summation of the vertical components of the effects of compensation beneath each unit area in the region surrounding the station, extending outward to the limit of appreciable effect. This is equally true whether there is local or regional compensation. The center of the compensation effect is necessarily at a material distance below sea level, and its vertical component will be large or near unity for some distance horizontally from the station. It is evident, therefore, that the compensation directly beneath the station itself will be only a small part of the total compensation affecting gravity at the station,[2] and that with this method the elevations for some distance about the station must be considered in estimating the thickness of the compensation plate. It is found that taking for the thickness of the compensation plate the average elevation within 100 miles of the station results in fairly close values for relative gravity.[3] The correctness of this computation of the compensation is not affected by any assumption as to local or regional compensation, or as to the thickness or density arrangement of the compensation, so long as the density is uniform horizontally, and the mean compensation effect is at a depth below sea level which is considerable in comparison with surface elevations; the thickness and density of the compensation may be varied inversely without effect if the mass remains constant. The principal source of uncertainty in this reduction is due to the method of selection of the average elevation plain for computing the compensation. From the attractive properties of an extended horizontal plate, the above shows that there is an indefinite number of combinations of compensation thicknesses and densities that would equally satisfy the conditions. The formula for the average elevation reduction is given in the next section of this paper.

The practical proof of the validity of a method of reduction of gravity observations, and of the theory on which it is based, is the extent to which gravity anomalies are reduced or eliminated by its use. The average elevation reduction is based on isostasy, while the Bouguer reduction is based on rigidity for the earth's crust. A comparison of the anomalies with these two methods gives the following results: For 13 continental

stations in mountainous regions (elevations to over 14,000 feet) the average isostatic anomaly is only 6% of the Bouguer, that is, 0.012 dyne as compared with 0.191, and in every case the former is smaller. For 16 oceanic island stations the reduction of average anomaly is from 0.162 to 0.053, and for 16 continental stations not in mountainous regions the reduction is from 0.063 to 0.034. For 27 stations of these two groups the isostatic anomaly is the smaller, and for only 3 is it the larger. For 22 coast stations, many with inadequate data, the isostatic reduction has a small advantage. The larger anomalies with the Bouguer reduction are eliminated in all these groups.

The locations of these observation stations include a great variety of the features of the earth's surface, with extreme conditions, and this, with their sufficient geographical extent, renders them competent to give evidence on isostasy. The results are consistent in eliminating excessive individual anomalies, and in reducing the average size of the anomalies for various conditions, and therefore this method gives substantial proof that the earth's crust is in a condition of equilibrium.

Numerically, this proof of isostasy does not differ materially from that obtained 15 years later by the Hayford method. Strictly comparable data are available for 42 stations, and for these the average anomalies are: Hayford 0.024 dyne, average elevation 0.025 and Bouguer 0.104.[4] The extreme anomalies at high mountain stations were satisfactorily accounted for. Thus for Pike's Peak the anomalies are Bouguer −0.239 dyne, Hayford +0.019, average elevation +0.006; and for Mauna Kea, Bouguer +0.244, Hayford +0.183, average elevation +0.068. The improvement over the Hayford anomalies here shown is probably due to the local compensation errors in the latter, as explained in the preceding paper. This is also indicated at the unusual station Seattle, where the comparison is: Bouguer −0.135, Hayford −0.095, average elevation −0.079.

That this strong proof of isostasy is obtained by an approximate method is due to the fact that the errors resulting from neglect of curvature and from the use of an average compensation are much less than the errors due to neglect of isostasy, and that the former errors to a large extent disappear in dealing with averages or differences of anomalies, while the latter do not.

It is important to observe that this approximate reduction method is not dependent on any assumption as to local or regional compensation, or the closeness of either. The averaging of the elevation within 100 miles is simply a mathematical convenience to obtain an approximate value for the compensation, free from the great errors of previous methods, and is not a proof of any degree of regional compensation; erroneous inferences have been drawn from misapprehension on this point.

The Average Elevation Reduction of Gravity Observations.—The following

is the formula for this isostatic method of reducing to sea level observed values of the force of gravity, as above described.

$$dg = +\frac{2gH}{r} - \frac{2gH3\delta}{r4\Delta} + T + \frac{2gH_1 3\delta}{r4\Delta}$$

$$= + \frac{2gH}{r} + T + \frac{2gh3\delta}{r4\Delta}$$

where dg = correction in dynes to observed gravity.
g = gravity in dynes at sea level.
H = elevation in meters of station above sea level.
H_1 = average elevation in meters above sea level, of region about the station.
h = $H_1 - H$ meters.
r = radius of the earth in meters.
δ = density of matter lying above sea level.
Δ = mean density of the earth.
T = the topographic correction for the effect of the departures of the surface about the station from a plain. This correction is appreciable at only a few stations, such as mountain summits; the computation is simple and is explained in Coast and Geodetic Survey, Appendix 1, 1894, pages 22, 23, and elsewhere; it may readily be made as exact as required.

In the second member of the first equation above, the first term $\left(\frac{3gH}{r}\right)$ is the correction for elevation, the second term $\left(\frac{2gH3\delta}{r4\Delta}\right)$ is the attraction of matter above sea level, and the last term $\left(\frac{2gH_1 3\delta}{r4\Delta}\right)$ is the compensation. In the second equation the last term $\left(\frac{2gh3\delta}{r4\Delta}\right)$ is the combined attraction and compensation;[5] for the accepted constants in meters and for average surface density (2.56), this becomes simply 0.000106 multiplied by the difference in meters between the average elevation in meters within 100 miles and the station elevation. Stating this in simple form, it becomes:

Reduction to sea level = Elevation correction + topographic correction + 0.000106 × (average elevation − station elevation, in meters).

The computation is simple. The average elevation within a radius of 100 miles from the station is read from maps by dividing the area into

convenient sized zones and compartments, allowing where necessary for the less density of sea water.[6]

American Contributions to Isostasy from 1889 to 1896.—Notable evidence indicating the existence of a condition of equilibrium in the crust of the earth had been given 50 to 80 years ago, by plumb line deflections, and by gravity observations,[7] in the Trigonometrical Survey of India, and the idea of isostasy had then been definitely proposed by English scientists, in explanation of these observed phenomena.

In the United States in 1889, Dutton published an important paper supporting the idea of isostasy. Shortly thereafter, in 1891, there was developed by the Coast and Geodetic Survey for measuring the relative force of gravity a pendulum apparatus which, compared with anything before it, was extremely accurate and efficient, and which made possible the field investigations of isostasy on the scale since carried out in this and neighboring countries.[8]

In 1894 gravity observations were carried out with this apparatus, which in their bearing on isostasy are the most effective single series that has been made. This one season's observations furnished the more significant part of the data used in the first developments of the system of reduction and proof of isostasy by the Hayford method. This work was planned by the Chief of the Coast and Geodetic Survey, Mendenhall, and the senior geologist of the Geological Survey, Gilbert.[9] This was only a few years after Dutton's paper, and its influence is seen in the plan. This series extended from the Atlantic Coast to the western plateaus, with stations situated so as to bring out the effect of all types of continental features, mountain summits, depressions, plateaus, river valleys and coasts; it included Pike's Peak, the most elevated summit where gravity has been measured.

In 1895 the Coast and Geodetic Survey published a report of this field work,[10] in which the writer included some discussion of the results of these and other well-distributed gravity observations, and their bearing on theories of the condition of the earth's crust. The method of reduction then developed and the numerical results then given are the same as those stated more systematically in the first parts of the present paper. The method involves approximations, and there were admitted deficiencies in its original presentation. But the tests were applied to the most extreme conditions found on the earth's land surface, continental and oceanic, and this reduction method on a basis of isostasy successfully met these tests, and eliminated or greatly reduced the anomalies of gravity, which all preceding methods had failed to do. The results were the first consistent proof of isostasy. The writer's last statement on this subject, published in that period (1896), showed his definite conclusion in favor of isostasy by referring to these results as "further confirming the validity

of the reduction to the sea level tentatively employed, and the theory of the condition of the earth's crust on which that reduction is based, the equilibrium or isostatic theory. This result has been obtained by combining observations made at altitudes above sea level ranging from 6 feet to 14,085 feet, and in a great variety of continental locations."[11]

Summary.—The large anomalies in the early gravity reductions were due to neglect of isostasy (Bouguer) and to seriously erroneous compensation (free-air method). These discrepancies were many times greater than would be accounted for by the small errors due to neglect of the effect of curvature of the earth's surface.

Definite proof of isostasy is furnished by a simple gravity reduction, using a reasonable average compensation.

This average elevation method is described. It affords a simple means of reducing gravity observations, with results, which though approximate, are in fair accord with the observed phenomena.

This method brings out certain conceptions of isostatic compensation, which are otherwise not so apparent. This reduction does not require any assumption as to the thickness or vertical density arrangement of the compensation, providing the mass of the compensation is adequate and its mean effect is at a considerable depth; therefore, an indefinite number of combinations of these elements will equally fit the data.

The resulting anomalies will be less dependable for individual stations than are given by a more complete analysis, but for high mountain summit stations they appear to be nearer the truth than local compensation anomalies.

These results alone do not give proof of the degree of local or regional compensation, but the inference reasonably drawn from them of the regional support of local features of the order of a single mountain is nearer the truth than the Hayford hypothesis of perfect local compensation. The writer's views as to the truth of isostasy and its degree of completeness, supported by the evidence given in the preceding paper are the same as those he reached in 1895.

Notable observational evidence of crustal equilibrium was developed in India the middle of the last century, and in consequence the idea of isostasy was clearly expressed by English scientists over 70 years ago.

Over a decade before the extensive and most important investigations of isostasy begun in 1906, there were definite pioneer contributions, in the United States, in observations, computations and deductions, supporting the equilibrium theory of the earth's crust. In 1894 was carried out the most effective single series of gravity observations that has been made, as respects the proof of isostasy. In 1895 was published a reduction method based on isostasy, and the results of a test made by applying this method to a considerable number of stations. This was the first gravity

reduction method and test which consistently eliminated the large anomalies.

This work by the writer in 1894 and 1895 was therefore the first consistent proof of the correctness of the equilibrium theory of the earth's crust, or isostasy.

[1] In this and following statements the effect of curvature is to be understood as including the effect of curvature on the attraction of distant topography and compensation; distant topography was not ignored in the earlier methods, but it was treated as though the sea level were a plain surface.

[2] Neglect of this fact causes the failure of the so-called free-air reduction. This reduction (see Fig. 2) is based on isostasy, as are all methods excepting the Bouguer, and does not neglect attraction or compensation, but it assumes that they exactly offset each other. In mountainous regions this would require that the compensation should affect gravity only at the station above it. This is physically impossible and, therefore, the free-air method cannot possibly account for the observed phenomena in uneven regions; in such regions it gives extreme anomalies, much larger even than those for a rigid earth. This reduction is, however, still in authoritative use.

The large free-air anomalies for summit and valley stations are eliminated by the average elevation reduction; for example, the free-air and average elevation anomalies are, respectively, for Pike's Peak +0.226 and +0.006, for Green River −0.090 and −0.003, and for Mauna Kea +0.628 and +0.068 dyne.

[3] Evidence from more complete methods of reduction, as given in the preceding paper, indicates that some radius near 100 miles represents the probable extent of regional compensation for high mountains, but this radius had no special significance in this average elevation reduction, and possibly a preferable method of estimating the average level could be found.

[4] For comparative data see: *Bulletin, Geological Society America*, **33**, 295, 296 (1922); *Journal, Washington Academy of Sciences*, **16**, 285–291 (1926). For method of proof, compare *Coast and Geodetic Survey, Special Publication*, No. **10** (1912).

[5] This was the form used in 1895. This term was then designated "lack of compensation," on the conception that as the effect of the attraction and the compensation almost exactly balance each other in a level region, a similar condition may be considered as extending through mountainous regions, leaving the topography above or the depression below the average plain as uncompensated locally, the uncompensated "local topographical irregularities" being of the order of "the attraction of a mountain on a station at its summit, or of an isolated island."

[6] This reduction was first used by the writer, and was rather inadequately described in *Coast and Geodetic Survey Report*, 1894, Appendix No. 1. It had not before been developed or used, but it was based on a long-neglected suggestion by Faye. It was briefly mentioned in an article in *Bulletin, Geological Society America*, **33**, 293 (1922). It is given under "more exact methods" by Swick in International Critical Tables, I, 1926, 402. Gilbert independently used this method in 1895, but with too small a radius, obtaining materially larger anomalies.

The resulting anomalies will be less dependable for individual stations than are given by a more complete analysis; but for high mountain stations they appear to be more nearly correct than local compensation anomalies.

[7] Lest this notable work be overlooked, the following account of it is quoted: "The Indian survey has also made valuable contributions to geodesy and geognosy in an elaborate series of pendulum observations for determining variations of gravity, which throws light both on the grand variation from the poles to the equator that governs the

ellipticity, and on the local and irregular variations depending on the constitution of the interior of the earth's crust. They were commenced in 1865 by Captain J. P. Basevi, on the recommendation of General Sabine and the Council of the Royal Society.* * * Captain Basevi had nearly completed the operations in India, and had taken swings at a number of stations of the Great Arc and at various other points near mountain ranges and coast lines, when he died of exposure in 1871 at a station on the highlands of the Himalayas, while investigating the force of gravity under mountain ranges." This work was completed by other officers, and the same pendulums were turned over to the United States Coast Survey, by whom they were swung at San Francisco and in Australia and Japan. The relation of the gravity and the plumb line results is discussed and it is stated that the pendulum observations revealed two facts: "First, that the force of gravity diminishes as the mountains are approached, and is very much less on the summit of the highly elevated Himalayan table-lands than can be accounted for otherwise than by a deficiency of matter below; secondly, that it increases as the ocean is approached, and is greater on islands than can be accounted for otherwise than by an excess of matter below." *Nature*, **32**, 485, 486 (1885).

[8] Swick, *Coast and Geodetic Survey, Special Publication*, No. **69** (1921), mentions those who took part in the development of this apparatus and refers to its importance to gravity work in this country, and adds: "The apparatus in use is practically the same as when first designed, * * * a tribute to the care and skill used in designing."

The writer had small part in the design, but contributed to the development of the use of this apparatus.

[9] The writer carried out these gravity observations, but the stations were selected before his connection with the work.

[10] *Report, Coast and Geodetic Survey*, **1894,** Appendix No. 1. A few additional results were given in *American Journal Science*, **1**, 186, 192 (1896).

The results were somewhat obscured by their publication incidental to a mass of field observations. Such work was not then under the Geodetic Division of the Coast Survey; this was an individual effort, and, of course, is not comparable in systematic development with the later work of a trained mathematical corps. The writer at the time took the results to indicate the support of local features of the order of a single mountain. This is probably near the truth, but these results alone really gave little evidence as between local and regional isostasy, and their bearing on the truth of isostasy is not affected by this question.

Hayford held an extreme view as to perfect local isostasy. The evidence is strongly against this view, and it appears to be no longer supported, but it undoubtedly influenced Hayford in his attitude toward the work of 1895, as did also the erroneous assumption that the average elevation method was based on a postulate of broad regional compensation. However, years later, in a letter to the writer, Hayford wrote: "It appears to me that what you did was to reach a close approximation, in 1895, to correct conclusions. * * * The fact that later and much more abundant evidence treated much more rigorously gives conclusions in such close agreement with those reached by you, emphasizes the validity of your work, and also strengthens the conclusions from the later work."

[11] *American Journal of Science*, **1**, 192 (1896).

SINGULAR POINTS OF VECTOR FIELDS UNDER GENERAL BOUNDARY CONDITIONS

BY MARSTON MORSE

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY

Communicated March 28, 1928

The proofs of the results in the following note have been completed and will be given at length in some other place.

Let $\sum_n$ be a closed manifold without boundary as defined by Veblen.[1] We require also that $\sum_n$ possess a Riemannian metric[2] defined by a positive quadratic form with coefficients of class C''. The requirement that $\sum_n$ have a Riemannian metric could be removed by suitable abstractions.

Let $(u) = (u_1, \ldots, u_r)$ be rectangular coördinates in an auxiliary Euclidian r-space. By an *r-element* on $\sum_n$, $0 < r \leqq n$, will be meant a set of points on $\sum_n$ homeomorphic with a simplex in the space (u). An r-element will be called *regular* if on the element the Riemannian coördinates (x) are functions of the u's of at least class C'', and if at least one set of r of the n x's has a jacobian with respect to the u's which is not zero. By a *regular r-manifold* on $\sum_n$ will be understood a non-singular closed r-manifold M_r on $\sum_n$, (Veblen, loc. cit.), each of whose points possesses a neighborhood on M_r which is a regular r-element. By a *regular r-manifold C_r with boundary B_{r-1}* we mean a connected sub-complex of a regular r-manifold bounded by a finite set B_{r-1} of distinct regular $(r-1)$-manifolds. By a *regular r-complex* we shall mean a finite set of distinct regular r-manifolds with or without boundary.

HYPOTHESIS α. *Let there be given on $\sum_n$ any regular n-manifold with boundary B_{n-1}. On B_{n-1} let there be given a continuous contravariant[2] vector field H_n which is non-singular on B_{n-1}.*

We define the index of an isolated singularity P of H_n as the index[3] of the corresponding field in the Euclidian space of the coördinates of a regular element containing P. The invariant properties of the later index have been enumerated by Hopf. The first theorem contains special hypotheses which will presently be removed.

THEOREM 1. *If the singularities of H_n are isolated, and if H_n is exteriorly normal to B_{n-1} at most at isolated points of a set s_{n-1} then the index-sum I_n of H_n is given as follows:*

$$I_n = (-1)^n E_n + I_{n-1}, \qquad n > 0$$

where E_n is the Euler-characteristic of C_n, and I_{n-1} is the index sum of the projection H_{n-1} of H_n on B_{n-1} at the points of s_{n-1}.

As a special case we may have s_{n-1} null and in this special case the

theorem has been proved by Hopf[4] and Lefschetz.[5] Lefschetz was concerned primarily with other matters and omitted the non-orientable case.

To remove the hypothesis that H_n be exteriorly normal to B_{n-1} at most at *isolated* points we need the following new notions.

Starting with the C_n and H_n of hypothesis α consider a set of regular complexes,

$$C_n,\ C_{n-1},\ \ldots,\ C_0 \tag{1}$$

of which C_r is either on the boundary B_r of C_{r+1} or null, $n > r \geqq 0$. Let H_r be the projection on B_r of H_{r+1}, null if B_r is null. Let σ_r and s_r be, respectively, the sets of points on the complexes B_r at which the vectors H_{r+1} are interiorly or exteriorly normal to B_r. Suppose, now, that the complexes (1) can be successively determined so that C_r contains all of the points s_r (in its interior if bounded), and none of the points σ_r. A set (1) so determined will be called a *topological set associated with* H_n.

That topological sets associated with the field H_n exist is intuitively obvious in all cases. *In the analytic case a particular choice of a topological set can always be defined by simple equations.*

THEOREM 2. *If with the vector field H_n on C_n there be associated a topological set* (1), *and if the singularities of H_n are isolated then*

$$I_n = E_0 - E_1 + \ldots + (-1)^n E_n \tag{2}$$

where I_n is the index sum of H_n, and E_i is the Euler characteristic of C_i.

This theorem is much more general than the first in that on the boundary H_n may be normal to C_n at arbitrary sets of points. To remove the restriction that the singularities of H_n be isolated we introduce a generalized index of infinite sets of singularities as follows.

Let the right-hand member of (2) be called the *alternating characteristic* of the topological set (1). Let s be any set of singularities of H_n which all lie on a complex $\overline{C}_n$ on C_n upon which there are no other singularities than those of s. If H_n on $\overline{C}_n$ possesses an associated topological set we define the generalized index of s to be the corresponding *alternating characteristic*. We show that the *generalized index is independent of the associated topological set used to define it.* The most general of our theorems is the following.

THEOREM 3. *If with a vector field H_n on C_n there be associated a topological set* (1), *the generalized index-sum of the field equals the alternating characteristic of the topological set of H_n.*

It has already doubtless been noted that the *alternating characteristic is a topological invariant of the complexes in a topological set.* To this we add that it is in general possible to so associate a topological set with an analytic field that a homeomorphism which is representable in the small

by a reversible analytic transformation will carry the field and associated topological set into an analytic field and associated topological set.

The application of these results to deformations, and coincidences of transformations of the same class does not need elaboration.

[1] Veblen, "Analysis Situs," *The Cambridge Colloquium*, Part II, § 24.

[2] Eisenhart, "Riemannian Geometry," *Princeton University Press*.

[3] Hopf, "Über die Curvatura integra geschlossener Hyperflächen," *Math. Ann.*, **95**, (1925), 340–351; Brouwer, "Über Abbildung von Mannigfaltigkeiten," *Ibid.*, **71**, (1911), 97–115.

[4] Hopf, "Vectorfelder in n-dimensionalen Mannigfaltigkeiten," *Ibid.*, **96** (1926), 225–249.

[5] Lefschetz, "Manifolds with a Boundary and Their Transformations," *Trans. Amer. Math. Soc.*, **29** (1927), 429–463.

REMARK ON THE NUMBER OF CLASSES OF BINARY QUADRATIC FORMS OF A GIVEN NEGATIVE DETERMINANT

By E. T. Bell

Department of Mathematics, California Institute of Technology

Communicated March 31, 1928

On p. 254 of Mathews' "Theory of Numbers," Part I, 1892 (all that was published), we find the following clear statement of a desideratum that has often been expressed. ". . . leads to the conclusion that in the series 1, 2, 3, . . . $(p-1)/2$ (p is an odd prime), there are more quadratic residues of p than non-residues. It does not appear that any independent proof of this proposition has ever been discovered. If any such proof could be found, it is not impossible that it might lead to a determination of h (the number of classes described in the title of this note) without the use of infinite series. Similar remarks apply to the other formulae for negative determinants."

The material for at least four such required proofs is contained in papers published by the writer during the past ten years. As the implied proofs are buried as mere details in other work, it is of interest to disinter one. On p. 113 of the *Tôhoku Mathematical Journal*, vol. **19**, Nos. 1, 2 (May, 1921), will be found the sufficient means for constructing a proof of the theorem depending only upon the rudiments of elementary algebra. In the paper cited, free use is made of elliptic theta identities. This is not necessary. In fact, as is evident, and is indeed well known, all basic identities of the kind used in the paper are obtainable from strictly elementary considerations; they have been so obtained by J. V. Ouspensky in a series of memoirs in the *Bulletin de l'Académie des Sciences de l'URRS*,

1925. Anyone who is interested may easily reconstruct the somewhat tedious entirely elementary proof for himself. I have constructed such a proof in detail. It is not worth reproduction. The thing can be done, but it does not seem worth while to do it.

The hope expressed by Mathews that h may be determined without the use of infinite series can be artificially realized on this basis. The gain in clarity by such a procedure is at best doubtful, and the like applies to the indicated proof concerning the distribution of quadratic residues. It seems rather futile to recast what is actually a matter of simple algebra in terms of so-called pure arithmetic when the latter is no more than a sophisticated disguise of the former.

CONCERNING CERTAIN TYPES OF NON-CUT POINTS, WITH AN APPLICATION TO CONTINUOUS CURVES[1]

BY HARRY MERRILL GEHMAN

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY

Communicated April 5, 1928

In our study of irreducible continua, we have been led to consider certain special types of non-cut points of a continuum. As these special types may be of use also in other connections, it seems best to set down here some of their properties, especially since an application is made in this paper to continuous curves, and in the following paper to irreducible continua.

In the following definitions, M is a continuum in space of any finite number of dimensions, P is a point of M, and D is a subset of M. If $M-P$ is connected, P is said to be a *non-cut point* of M; if $M-P$ is not connected, P is said to be a *cut point* of M. The set $M-D$ is said to be the complement of D. If the complement of D is closed, D is said to be an *open subset* of M. If D is a connected open subset of M, then D is said to be a *domain* in M, or an *M-domain*.

We shall say that a point P of M is a *point of type 1*, if given any positive number ϵ there exists an M-domain containing P of diameter less than ϵ whose complement is connected. We shall say that P is a *point of type 2*, if given any positive number ϵ there exists an open set in M containing P of diameter less than ϵ whose complement is connected. For convenience, we shall call a non-cut point of M a *point of type 3*.

THEOREM 1. *If P is a point of type 1 of a continuum M, it is a point of type 2, but not conversely. If P is a point of type 2, it is a point of type 3, but not conversely.*

THEOREM 2. *A point P of type 2 of a continuum M has the property that if N is any closed subset of $M-P$, then $M-P$ contains a continuum containing N. Conversely, any point with this property is a point of type 2.*

THEOREM 3. *If P is a point of type 1 of a continuum M, then M is connected im kleinen[2] at P.*

Proof.—The condition that M be connected im kleinen at P is logically equivalent to the condition that M contain an M-domain containing P of diameter less than an arbitrary positive number ϵ, which proves the theorem.

The following example shows that if M is connected im kleinen at P, and P is a point of type 3, it does not necessarily follow that P is a point of type 1, or even of type 2. See, however, theorem 4.

Example.—Let the continuum M consist of all points of the straight line intervals from (0,0) to (1,0); from (0,0) to (0,1); from $(1/n,0)$ to $(1/n,1)$, for $n = 1,2,3,\ldots$ The point $P = (0,0)$ is a point of type 3 of M, but not of type 1 or 2, and M is connected im kleinen at P.

THEOREM 4. *Every point of type 3 of a continuous curve (bounded or unbounded) is a point of type 1.* In other words: *If P is a non-cut point of a continuous curve M (bounded or unbounded) in space of any finite number of dimensions, then given any positive number ϵ, there exists an M-domain containing P of diameter less than ϵ whose complement is connected.*

Proof.—Let C denote an n-dimensional sphere with radius $\epsilon/4$ about P as center, and let E denote its exterior. The connected subsets of $M\cdot(C + E)$ that contain points at a distance greater than or equal to $\epsilon/2$ from P lie in a *finite* number of maximal connected sets, which we shall denote by $N_1, N_2, \ldots, N_k$. For if the number of these sets were infinite, M would fail to be connected im kleinen at any limit point of H, where H is a set of points of M at a distance equal to $\epsilon/2$ from P, no two points of H being selected from the same maximal connected subset of $M\cdot(C + E)$.

Let us select a point P_i from each of the sets N_i, and let us join each of the points $P_2, P_3, \ldots, P_k$ to P_1 by an arc in $M-P$.[3] Then the set $N = N_1 + N_2 + \ldots + N_k + P_1P_2 + P_1P_3 + \ldots + P_1P_k$ is a continuum which does not contain P, but contains all points of M whose distance from P is greater than or equal to $\epsilon/2$. If D denotes the maximal connected subset of $M-N$ that contains P, the set $M-D$ is connected and D is an M-domain[4] containing P of diameter less than ϵ.

THEOREM 5. *If P is a cut point of a continuous curve M (bounded or unbounded) in space of any finite number of dimensions, then given any positive number ϵ, there exists an M-domain containing P of diameter less than ϵ whose complement consists of a finite number of maximal connected sets.*

Proof.—The set $M-P$ contains only a finite number of maximal con-

nected subsets that have points at a distance greater than or equal to $\epsilon/2$ from P. If we add P to each of these sets, we obtain a finite collection $N_1, N_2, \ldots, N_k$ of subcontinua of M such that (1) P is a non-cut point of each of the sets, (2) $N_i \cdot N_j = P$, if $i \neq j$, and (3) every point of M at a distance from P greater than or equal to $\epsilon/2$ is a point of one of the sets N_i. By theorem 4, each of the sets N_i contains a domain D_i containing P of diameter less than $\epsilon/2$ whose complement is connected. The set $D = M - (N_1 - D_1) - (N_2 - D_2) - \ldots - (N_k - D_k)$ is connected, and its complement is closed, since it consists of the finite collection of continua $N_1 - D_1, \ldots, N_k - D_k$. The set D is therefore an M-domain containing P. The diameter of D is less than ϵ, because every point of D is at a distance less than $\epsilon/2$ from P.

It is interesting to note that if the method of proof employed in theorem 4 is applied to a *plane* continuous curve, the set $M - D$ so obtained is itself a continuous curve.[5] Similarly, in theorem 5, $M - D$ consists of a finite number of continuous curves.

[1] Presented to the American Mathematical Society, April 7, 1928.

[2] H. Hahn, *Wiener Berichte*, **123**, IIa, 1924 (2433–2489).

[3] R. L. Moore, *Bull. Amer. Math. Soc.*, **29**, 1923 (289–302), theorem A.

[4] R. L. Wilder, *Fund. Math.*, **7**, 1925 (340–377), theorem 9.

[5] H. M. Gehman, *Annals Math.*, **28**, 1927 (103–111), theorems 6 and 8.

CONCERNING IRREDUCIBLE CONTINUA[1]

BY HARRY MERRILL GEHMAN

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY

Communicated April 5, 1928

A point set M is said to be *an irreducible continuum about a point set* A, if M is a continuum and contains A, but contains no proper subset which is a continuum and contains A. In a previous paper,[2] we have made a brief study of irreducible continua, and we have noted there that a bounded continuum M may be an irreducible continuum about a point set A which contains no non-cut points of M. This differs from the case where M is irreducibly connected[2] about a set A, because in that case A contains all the non-cut points of M. The object of the present paper is to show that nevertheless certain relations *do* exist between the set A and the set of non-cut points of M, in case M is an irreducible continuum about A. A number of previously known theorems are obtained as corollaries to the theorems of this paper.

All the theorems hold true for continua in space of any finite number of dimensions. We shall use *continuum* here in the sense of a closed and

THEOREM 2. *A point P of type 2 of a continuum M has the property that if N is any closed subset of $M-P$, then $M-P$ contains a continuum containing N. Conversely, any point with this property is a point of type 2.*

THEOREM 3. *If P is a point of type 1 of a continuum M, then M is connected im kleinen*[2] *at P.*

Proof.—The condition that M be connected im kleinen at P is logically equivalent to the condition that M contain an M-domain containing P of diameter less than an arbitrary positive number ϵ, which proves the theorem.

The following example shows that if M is connected im kleinen at P, and P is a point of type 3, it does not necessarily follow that P is a point of type 1, or even of type 2. See, however, theorem 4.

Example.—Let the continuum M consist of all points of the straight line intervals from (0,0) to (1,0); from (0,0) to (0,1); from $(1/n,0)$ to $(1/n,1)$, for $n = 1,2,3,\ldots$ The point $P = (0,0)$ is a point of type 3 of M, but not of type 1 or 2, and M is connected im kleinen at P.

THEOREM 4. *Every point of type 3 of a continuous curve (bounded or unbounded) is a point of type 1.* In other words: *If P is a non-cut point of a continuous curve M (bounded or unbounded) in space of any finite number of dimensions, then given any positive number ϵ, there exists an M-domain containing P of diameter less than ϵ whose complement is connected.*

Proof.—Let C denote an n-dimensional sphere with radius $\epsilon/4$ about P as center, and let E denote its exterior. The connected subsets of $M\cdot(C + E)$ that contain points at a distance greater than or equal to $\epsilon/2$ from P lie in a *finite* number of maximal connected sets, which we shall denote by $N_1, N_2, \ldots, N_k$. For if the number of these sets were infinite, M would fail to be connected im kleinen at any limit point of H, where H is a set of points of M at a distance equal to $\epsilon/2$ from P, no two points of H being selected from the same maximal connected subset of $M\cdot(C + E)$.

Let us select a point P_i from each of the sets N_i, and let us join each of the points $P_2, P_3, \ldots, P_k$ to P_1 by an arc in $M-P$.[3] Then the set $N = N_1 + N_2 + \ldots + N_k + P_1P_2 + P_1P_3 + \ldots + P_1P_k$ is a continuum which does not contain P, but contains all points of M whose distance from P is greater than or equal to $\epsilon/2$. If D denotes the maximal connected subset of $M-N$ that contains P, the set $M-D$ is connected and D is an M-domain[4] containing P of diameter less than ϵ.

THEOREM 5. *If P is a cut point of a continuous curve M (bounded or unbounded) in space of any finite number of dimensions, then given any positive number ϵ, there exists an M-domain containing P of diameter less than ϵ whose complement consists of a finite number of maximal connected sets.*

Proof.—The set $M-P$ contains only a finite number of maximal con-

nected subsets that have points at a distance greater than or equal to $\epsilon/2$ from P. If we add P to each of these sets, we obtain a finite collection $N_1, N_2, \ldots, N_k$ of subcontinua of M such that (1) P is a non-cut point of each of the sets, (2) $N_i \cdot N_j = P$, if $i \neq j$, and (3) every point of M at a distance from P greater than or equal to $\epsilon/2$ is a point of one of the sets N_i. By theorem 4, each of the sets N_i contains a domain D_i containing P of diameter less than $\epsilon/2$ whose complement is connected. The set $D = M - (N_1 - D_1) - (N_2 - D_2) - \ldots - (N_k - D_k)$ is connected, and its complement is closed, since it consists of the finite collection of continua $N_1 - D_1, \ldots, N_k - D_k$. The set D is therefore an M-domain containing P. The diameter of D is less than ϵ, because every point of D is at a distance less than $\epsilon/2$ from P.

It is interesting to note that if the method of proof employed in theorem 4 is applied to a *plane* continuous curve, the set $M - D$ so obtained is itself a continuous curve.[5] Similarly, in theorem 5, $M - D$ consists of a finite number of continuous curves.

[1] Presented to the American Mathematical Society, April 7, 1928.

[2] H. Hahn, *Wiener Berichte*, **123**, IIa, 1924 (2433–2489).

[3] R. L. Moore, *Bull. Amer. Math. Soc.*, **29**, 1923 (289–302), theorem A.

[4] R. L. Wilder, *Fund. Math.*, **7**, 1925 (340–377), theorem 9.

[5] H. M. Gehman, *Annals Math.*, **28**, 1927 (103–111), theorems 6 and 8.

CONCERNING IRREDUCIBLE CONTINUA[1]

BY HARRY MERRILL GEHMAN

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY

Communicated April 5, 1928

A point set M is said to be *an irreducible continuum about a point set* A, if M is a continuum and contains A, but contains no proper subset which is a continuum and contains A. In a previous paper,[2] we have made a brief study of irreducible continua, and we have noted there that a bounded continuum M may be an irreducible continuum about a point set A which contains no non-cut points of M. This differs from the case where M is irreducibly connected[2] about a set A, because in that case A contains all the non-cut points of M. The object of the present paper is to show that nevertheless certain relations *do* exist between the set A and the set of non-cut points of M, in case M is an irreducible continuum about A. A number of previously known theorems are obtained as corollaries to the theorems of this paper.

All the theorems hold true for continua in space of any finite number of dimensions. We shall use *continuum* here in the sense of a closed and

connected point set *containing more than one point.* We shall use the following notation: If K is a point set, K' denotes the set consisting of K and all points which are limit points of K.

THEOREM 1. *If a bounded continuum M is an irreducible continuum about a set A, then some points of A' are non-cut points of M.*[3]

Proof.—Let M be a bounded continuum which is an irreducible continuum about a set A, every point of which is a cut point of M. Let $L = P_1 + P_2 + P_3 + \ldots$, be a countable subset of A such that $L' = A'$. Since P_1 is a cut point of M, $M = H_1 + K_1$, where H_1 and K_1 are continua, and $H_1 \cdot K_1 = P_1$. Let P_{i_2} be the first point of L in $H_1 - P_1$. Then $M = H_2 + K_2$, where H_2 and K_2 are continua, $H_2 \cdot K_2 = P_{i_2}$, K_1 is a proper subset of K_2, and H_2 is a proper subset of H_1. For each positive integer n, having defined H_n and K_n, let $P_{i_{n+1}}$ be the first point of L in $H_n - P_{i_n}$. (Note that $i_n < i_{n+1}$.) Then $M = H_{n+1} + K_{n+1}$, where H_{n+1} and K_{n+1} are continua, $H_{n+1} \cdot K_{n+1} = P_{i_{n+1}}$, K_n is a proper subset of K_{n+1}, and H_{n+1} is a proper subset of H_n. If at any step $L \cdot (H_n - P_{i_n}) = 0$, then K_n which is a proper subcontinuum of M, contains $L' = A'$, which is contrary to hypothesis. Therefore, this process can be continued indefinitely.

Let $H^* = H_1 \cdot H_2 \cdot H_3 \ldots$, and $K^* = K_1 + K_2 + K_3 + \ldots$. Evidently $M = H^* + K^*$, where $H^* \cdot K^* = 0$, K^* is connected, H^* is a continuum, and $H^* \cdot L = 0$. Since K^* contains L, every point of H^* is a limit point of K^*, otherwise a proper subcontinuum of M contains $L' = A'$. It follows that every point of H^* is a non-cut point of M.

Let P be a limit point of the set $G = P_1 + P_{i_2} + P_{i_3} + \ldots$ The point P is a point of each of the sets H_n, because for each value of n, the set K_n contains a finite number of points of G, and H_n contains an infinite number of points of G. The point P is therefore a point of H^* and is a non-cut point of M. Since P is also a point of A', this proves the theorem.[4]

THEOREM 2. *If a bounded continuum M is an irreducible continuum about a set A, then M is an irreducible continuum about the set consisting of those points of A' which are non-cut points of M.*

Proof.—Suppose on the contrary that M contains a proper subcontinuum N which contains all those points of A' which are non-cut points of M. Let P_{i_1} be the first point of L in $(M - N) \cdot A$. Then $M = H_1 + K_1$, where H_1 and K_1 are continua, $H_1 \cdot K_1 = P_{i_1}$, and K_1 contains N. Proceeding as in proof of theorem 1, we obtain the point P of A' which is a non-cut point of M, but is not a point of N. This contradicts our assumption concerning N.

COROLLARY 2*a*. *Every bounded continuum contains at least two non-cut points.*[5]

Proof.—Since M is an irreducible continuum about itself, M contains at least one non-cut point of itself, by theorem 1. Since no continuum is an irreducible continuum about a set consisting of a single point, it follows from theorem 2 that M contains at least two non-cut points.

COROLLARY 2*b*. *Every bounded continuum is an irreducible continuum about the set consisting of all its non-cut points.*[6]

THEOREM 3. *If a continuum (bounded or unbounded) is an irreducible continuum about a set A, then A' contains all points of type 2 of the continuum.*[7]

This theorem is a direct consequence of theorem 2 of the preceding paper. From this theorem and theorems 2 and 4 of the preceding paper follows:

THEOREM 4. *If M is a continuous curve (bounded or unbounded) in space of any finite number of dimensions, which is an irreducible continuum about a set A, then A' contains all the non-cut points of M.*[3]

[1] Presented to the American Mathematical Society at Nashville, Dec. 28, 1927.

[2] H. M. Gehman, these PROCEEDINGS, **12**, 1926 (544–547).

[3] For the case where M is a (bounded) plane continuous curve, see: H. M. Gehman, *Amer. J. Math.*, **49**, 1927 (189–196), theorem 3.

[4] The author wishes to thank Professor W. A. Wilson for suggesting certain simplifications in the proof of theorem 1.

[5] R. L. Moore, *Bull. Amer. Math. Soc.*, **29**, 1923 (289–302), theorem C.

[6] Gehman, these PROCEEDINGS, loc. cit., theorem 4.

[7] See preceding paper for definition of a point of type 2.

CRITERIA FOR THE SIMPLIFICATION OF ALGEBRAIC PLANE CURVES

BY JULIAN COOLIDGE

DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY

Communicated March 22, 1928

It is the purpose of the present note to establish necessary and sufficient conditions for the possibility of effecting certain notable simplifications of a given algebraic plane curve by means of a Cremona transformation. It will be recalled that every such transformation can be factored into the product of a finite number of quadratic transformations and colineations.

Given a curve of order n, which we may suppose irreducible, let it have singular points $P_1P_2 \ldots$ whose actual multiplicity is $r_1r_2 \ldots$, it being understood that these points may be either distinct or infinitely near in the sense commonly used. The theorems used and established are to be birationally invariant, so it is immaterial how near the singular points lie to one another. A curve of order $n-3$ which has a singular

point of order r_i-1 or more wherever the original curve has multiplicity r_i is called a *special adjoint of index* 1. If it be of order $n-3k$ with multiplicity at least r_i-k where the original curve has multiplicity r_i, or multiplicity at least 0 when this number is negative, it is called a *special adjoint of index k*. If the system of all such curves be infinite, there may be variable curves, and fixed curves, the general curve being composed of all the fixed curves, perhaps counted several times each, and one variable curve. Such fixed curves are said to be *fundamental*, they have no non-singular intersections with the given curve.

The following theorem, which is now classic,[1] is the basis of all that follows:

If an algebraic plane curve be subjected to a Cremona transformation, the system of special adjoints of each index either goes over into the corresponding system or a part thereof, the rest being a fundamental curve or fundamental curves corresponding to the singular points of the transformation.

When we are dealing with a quadratic transformation we may introduce or abolish one or more of the sides of the fundamental triangle counted a number of times. The number of new lines introduced is the number of vertices where the special curve has multiplicity 0.

Suppose that the system of special adjoints of index k have the property that they are merely required to have a certain order, namely, $n-3k$, then the original curve can have no singular point of multiplicity greater than k. Let us find the necessary and sufficient conditions for the attainability of this:

THEOREM 1. *Given a curve whose special adjoint system of index k has the following properties:*

(1) The fundamental curves are rational;

(2) The base points of the adjoint system on the given curve comprise all the singularities of each fundamental curve and form a normal base, determining it uniquely;

(3) The total number of such base points is equal to the number of different fundamental curves;

(4) Through each base point on the given curve will pass at least two fundamental curves: then

The given curve can be transformed into one having no singularity of order higher than k.

Usually there are no base points for the adjoints other than the assigned singularities of the given curve, but such might conceivably arise. Now to prove the theorem. Suppose that we have a curve of order ν with singular points of orders $\rho_1\rho_2 \ldots$ which impose independent conditions on the curve. Then an immediate calculation shows that the expression

$$\nu(\nu + 3) - \Sigma\rho_i(\rho_i + 1)$$

is an invariant for the curve under all quadratic transformations, provided

the summation be extended to include any new singular points introduced by the transformation. In the case of a rational curve uniquely determined by a normal base including all of its singularities we have

$$\begin{gathered} \nu(\nu + 3) - \Sigma\rho_i(\rho_i + 1) = 0 \\ \nu(\nu - 3) - \Sigma\rho_i(\rho_i - 1) = -2 \\ 3\nu - 1 - \Sigma\rho_i = 0 \\ \nu(\nu - 3l) = \Sigma\rho_i(\rho_i - l) - (l - 1). \end{gathered}$$

This equation seems to tell us that the total number of intersections of the fundamental curve with a special adjoint of index 1 is less than the number at the singular points alone. Hence there can be no special adjoints of index greater than 1, nor of index 1 either, for the curve is rational by (1). Now it is known that if a curve be rational, the N.S. condition that it can be carried into a straight line by means of a Cremona transformation is that all special adjoint system become impossible.[2] The fundamental points of the quadratic transformations are singular points of this curve, and so by (2) of the curve of order n. Hence we may reduce one fundamental curve to be of order unity without introducing any new ones. This curve, or rather straight line, is determined by a normal base, two points. Through each will pass at least one other fundamental curve. One of these, if not a line, will have at least two other base points, for a curve or order above two which is rational cannot be determined by just two points. Hence we have a base point off of our line. Using this and the two base points on the line as fundamental in a quadratic transformation, we abolish the line as a fundamental curve of the system of index k, yet without introducing any new fundamental curve. Continuing thus we abolish all the fundamental curves, hence, by (3) we abolish all the base points. Then the system consists in the totality of curves of order $n-3k$, and the original curve had no singularity of order above k.

Suppose that, in particular, $k = 1$. Then we can transform our original curve to one with no singular points. Conversely, suppose we have a non-singular curve. If the special adjoint system of index 1 do not exist, it is rational and a line or conic, and none of the subsequent systems exist. If it be not rational, these conditions are fulfilled.

THEOREM 2. *The necessary and sufficient condition that it be possible to transform a curve by a Cremona transformation into one with no singular points is that either it is rational with no special adjoints of any index, or that the conditions of theorem 1 are fulfilled for the adjoint system of index 1.*

Take $k = 2$. Then we can transform to a curve with no singular points but double points. Conversely, suppose we have a curve whose only singular points are double points, distinct or infinitely near. Then if the adjoint system of index 2 exists, these conditions are satisfied. If the

system of index two does not exist, but that of index 1 does, then the curve is of order five or less, for a system of curves of order 0 with no assigned points may be said to exist. If of order 5 it is not hyperelliptic. Again, if the fundamental curve of the system of index 1 be a line or a conic or nothing at all, and if the curve be not hyperelliptic of deficiency 3, the curve is of order 5 or less but not a quintic with just a triple point, and it is easy to see that it can be carried into a curve with only singular points of the second order.

THEOREM 3. *The necessary and sufficient condition that it be possible to transform a curve by a Cremona transformation into one with no singular points but double points, or with none at all, is that either it is rational lacking all special adjoint systems, or that it lacks the system of index 2 but that of index 1 can be transformed to a line or conic or nothing at all, yet the curve is not hyperelliptic of deficiency 3, or, lastly, the system of index 2 exists and obeys the conditions of theorem 1.*

It is well known that any curve can be carried on a Cremona transformation into one whose only singular points are ordinary ones, and by a birational transformation, that is to say, a transformation birational for the curve alone, into one whose only singular points are nodes. Our theorem 3 shows that, in general, this cannot be accomplished by a Cremona transformation. I do not know what additional restrictions beside those in 3 are necessary in order that the double points arrived at should all be distinct with distinct tangents.

[1] Conf. Enriques-Chesini: "Teoria geometrica delle equazioni," Vol. III, Bologna, 1924, pp. 178 and 187.

[2] *Ibid.*, p. 188.